Report to Congress:

DRG Refinement: Outliers, Severity of Illness, and Intensity of Care

Department of Health and Human [illegible]s
Health Care Financing Administration
Office of Research and Demonstrations

PUBS
RA
971
.3
J451
1987

RA
971.3
.J451
1987

Report to Congress:

DRG Refinement: Outliers, Severity of Illness, and Intensity of Care

Department of Health and Human Services
Health Care Financing Administration
Office of Research and Demonstrations

HCFA Pub. No. 03254

ACKNOWLEDGMENTS

This report was prepared by the Office of Research, Health Care Financing Administration. The research for this report was managed by and this report was written by Stephen F. Jencks, M.D., under the supervision of Allen Dobson, Ph.D. Terrence Kay and Julie Walton provided technical editorial assistance, and Joanna Bandzwolek and Jacqueline Weigman provide production assistance.

Shan Cretin, Joseph P. Newhouse, Linda Worthman, Kenneth E. Thorpe, Grace M. Carter, Paul B. Ginsburg, Emmett B. Keeler, Joan Buchanan, and Christina Witsberger at the Rand Corporation; Janet B. Mitchell, Jerry Cromwell, Kathleen A. Calore, Brooke Harrow, and Helene Hewes at Health Economics Research; Stuart Guterman, Stephen F. Jencks, Terrence Kay, John F. Langenbrunner, and Sherry Terrell at Health Care Financing Administration; and Richard Averill at Health Services International all conducted special research studies or wrote special papers in support of this report. Those papers are marked with an asterisk in the reference list.

TABLE OF CONTENTS

Executive Summary and Introduction Page

Chapter 1. The Goals and Methods of the PPS Case Mix Adjustment System 1 - 1

I. The Goals of Case Mix Adjustment 1 - 2
II. The Structure of the Current System 1 - 7
III. Equity and Payment Discrepancies 1 -12

Chapter 2. Predicting Needed Care For Individual Patients 2 - 1

I. The PPS Model of Differences in Treatment Needs 2 - 1
II. Defining Necessary Care 2 - 3
III. Predictors of the Cost of Necessary Care: The Independent Variables 2 -13
IV. Evaluating DRGs 2 -19

Chapter 3. Evaluation of Alternative Case Mix Measures 3 - 1

I. Criteria for Different System Uses 3 - 1
II. Acute Physiology and Chronic Health Evaluation (APACHE) 3 - 2
III. Disease Staging (Staging) 3 - 7
IV. Medical Illness Severity Grouping System (MEDISGRPS) 3 -12
V. Patient Management Categories (PMCs) 3 -15
VI. Severity of Illness Index (SOII) and Computerized Severity Index (CSI) 3 -19
VII. The Validity of Cost Predictors 3 -26
VIII. The Feasibility of Alternative Case Mix Systems for PPS 3 -32

Chapter 4. How Do Costs of Treatment Differ Between Hospitals? 4 - 1

I. The PPS Model of Interhospital Case Mix Differences 4 - 2
II. How Patients with High Treatment Needs Concentrate in Some Hospitals 4 -10
III. Empirical Evidence on the Differentiation of Hospitals 4 -14
IV. Setting DRG Weights and the Problem of Comparing Hospitals 4 -20
V. The "Accuracy" of Payments in the First Year of PPS 4 -25
VI. Outliers Payments and Equity to Hospitals 4 -38
VII. Should PPS Pay According to Hospital Size? 4 -50
VIII. Hospital-Level Variables for Case Mix Adjustment 4 -52
IX. The Adequacy of Current Case Mix Adjustment For National Rates 4 -54

Chapter 5. Research for Improving Case Mix Measurement Under PPS 5 - 1

I. Improving Case Mix Measurement at the Individual Case Level 5
II. Improve Payment of Outlier Cases 5 3
III. Improving Adjustments at the Hospital Level 5 - 4
IV. Timetable 5 - 4

EXECUTIVE SUMMARY
AND INTRODUCTION

I. BACKGROUND

A. CONGRESSIONAL MANDATE. In the Social Security Amendments of 1983, Congress instructed the Secretary of the Department of Health and Human Services to report on "the appropriateness of the factors used...to compensate hospitals for the additional expenses of outlier cases, and the application of severity of illness, intensity of care, or other modifications to the diagnosis-related groups, and the advisability and feasibility of providing for such modifications." (P.L. 98-21, section 603(a)(2)(C)(iii)). This report establishes a conceptual framework for considering these issues, provides a summary of research conducted to answer the question, and presents options for future research.

B. OUTLIERS, SEVERITY, INTENSITY, AND THE PROSPECTIVE PAYMENT SYSTEM. Medicare's Prospective Payment System (PPS) controls Medicare outlays and provides incentives for efficiency by determining in advance the amount which Medicare will pay a hospital for a hospitalization or "case". Because cost per case differs among disorders and among patients with similar disorders, payments to each hosptial must be adjusted to reflect the "case mix" of that particular hospital. The accuracy of this process of case mix adjustment is critical to the equity and long term success of PPS.

When the Administration and Congress designed PPS, research had shown that five variables had a measurable influence on how a hospital's average cost per admission compared with that of other hospitals. The designers of PPS constructed the system so that it would determine payments based on four of these variables:

-- Case Mix Index (CMI)(the average resource requirements of a case as measured by diagnosis-related groups(DRGs)),

-- Labor Costs (measured by the Area Wage Index),

-- Urban/Rural Location (measured by the size of the Metropolitan Statistical Area, if any, in which the hospital is located), and

-- Teaching Intensity (measured by the number of interns and residents per hospital bed).

The designers of PPS knew that a fifth variable,

-- Hospital size (measured by number of beds)

also predicted variation in cost per case, but they decided against determining payments on the basis of hospital size.

Table ES.1A shows that these five factors can explain about 69-72 percent of variation in average operating cost per case among hospitals. As noted, PPS does not consider hospital size in determining payments; Table ES.1B shows that case mix index, labor costs, urban/rural location, and teaching intensity without hospital size explain 66-69 percent of variation. PPS does not, for reasons explained in the report (Chapter 4.I) use case mix index, labor costs, urban/rural location, and teaching intensity in a way which most accurately simulates current costs; Table ES.1C shows that when these factors are used in the way that PPS was actually implemented, they explain 64-66 percent of variation. The fact that the payment system accounts for only two thirds of variation raises a question whether the remaining third is due to factors for which Medicare should not pay or whether the system is imprecise. Four factors are thought to account for most of the remaining variation:

-- Outliers (cases so unusual that they were not properly described by the DRG in which they were classified).

-- Severity of Illness (defined as the variation in the clinical illness within a DRG).

-- Intensity of Care (defined as variation in care for patients in the same DRG with the same clinical severity of illness). Some variation in intensity of care is medically appropriate while some variation constitutes either wasteful or inappropriate care).

-- Efficiency (inefficiency is obvious when the cost of providing an individual service is unnecessarily high; inefficiency and intensity of care may be difficult to distinguish if the need for a service is questionable).

Satisfactory measures were not available for efficiency or for variations in severity and intensity within DRGs. Outliers could be identified, but estimates of their importance were not available. The design of PPS protected hospitals against these uncertainties by several cushioning devices, notably a transition period before hospitals would be paid wholly under a national DRG-based system. In addition, Congress asked for this report before the end of the transition period. Thus, the congressional mandate to report on outliers, severity, and intensity requires examining the unexplained variation in average cost per case among hospitals.

C. PPS: INCENTIVES, EQUITY, AND PROTECTION. We have already indicated that many payment factors other than the DRG and outlier systems provide case mix adjustment. The corrections for teaching activity, hospital-specific and regional costs, and urban/rural location are implicitly corrections for variations in severity of illness and intensity of care. Thus, we can only evaluate the DRG and outlier systems, and the importance of unmeasured variation of severity and intensity, as part of the overall payment system. Three concepts are important for this evaluation:

1. Incentives. The key incentives for efficiency in PPS are that payment is determined before services are rendered and that the unit for payment is the hospitalization, a large "bundle" of individual services. Available data indicate that such incentives are strong even if hospitals are

cushioned by various adjustments and that these incentives are almost independent of whether rates are national, regional, or even partly hospital-specific. The Secretary's Report on the Impact of the Medicare Prospective Payment System for 1984 described dramatic hospital responses to these incentives in the first year of PPS, including large decreases in length of stay, staffing, and costs. These responses make clear that the incentives were very effective even when payments were largely based on historical hospital costs.

2. Equity. Case mix adjustment must achieve two kinds of equity: it must pay hospitals an amount which covers the cost of efficiently providing necessary care of Medicare beneficiaries and it must pay in a way which does not give hospitals an incentive to discriminate against any class of beneficiaries or against beneficiaries with any particular clinical condition.

If PPS does not pay some hospitals on average enough to provide medically appropriate care, these hospitals may face unfair financial hardship and possible insolvency; beneficiaries may face decreased quality of care or difficulty in securing care. Unfairness will occur if outliers, severely ill cases, or more intense care concentrate in certain hospitals, in a way which is not otherwise recognized by the payment system.

This report focuses on the ability of the PPS case mix adjustment system to achieve both kinds of equity. A key feature of equity to hospitals is that payments need not be accurate for each admission, but they must be equitable on average. On the other hand, equity to patients requires that a hospital not be able to identify in advance that any patient will need more care than Medicare will pay for. Such beneficiaries would be targets for discrimination and inadequate care.

3. Protection. The teaching and urban/rural adjustments were designed to cushion urban and teaching hospitals against large changes in payments. To guard against possible inequities, Congress also:

- enacted provisions for special payments for outlier cases,
- doubled the Department's proposed adjustment for indirect costs of teaching from 5.79 percent to 11.58 percent,
- created a transition period during which hospital payments would be based in part on the hospital's historical costs and on regional rather than national payment rates,
- excluded certain types of psychiatric, rehabilitation, long-term, and children's facilities,
- provided for special treatment of cancer hospital, sole community hospitals, rural referral centers, and hospitals with a "disproportionate" share of poor and Medicare patients, and
- required this report on the appropriateness of the outlier payment system, as well as on the severity and intensity issues, before the transition period was scheduled to expire.
- Congress also inacted higher PPS updates than the Secretary deems necessary for efficient and quality health care.

Hospital-specific, regional, urban/rural, and teaching adjustments were all designed to cushion hospitals from undue dislocation as PPS was put in place. To the extent that any of these "cushions" is removed (and the law schedules hospital-specific and regional adjustments to phase out by fiscal year 1988), accurate case mix adjustment becomes increasingly critical to the long-term viability of PPS.

The key issue in continued refinement of PPS is preserving incentives without endangering either the system's equity or its viability.

C. PREDICTING COST PER CASE. Because the components of case mix adjustment are interdependent, compromises or imprecision in measuring or applying one component may influence the apparent accuracy of another. Several such compromises were made in the design of PPS. Although CMI, labor cost, urban/rural location, teaching intensity, and hospital size can explain 72 percent of variation in average cost per case, some compromises were made in the design of PPS that reduce its precision below that level (compare Tables 4.1A and 4.1C and see Chapter 4.I).

1. Labor Cost. Hospital cost per case rises almost exactly 1 percent for every 1 percent increase in the wage index, but PPS allows payments to change only 0.79 percent because labor costs contribute directly to only 79 percent of hospital costs.

2. Urban/Rural Location. The Department's original calculations (Pettengill and Vertrees 1982) used four levels of urbanicity (from large city to rural), but PPS distinguishes only between urban and rural locations in setting standard payment amounts. Moreover, PPS treats all rural areas in a State as a single area in setting wage adjustments.

3. Hospital Size. Research shows that average cost per case increases with hospital size, but conventional economic theory suggested that this increase was due to inefficiency. Paying more to large hospitals also appeared to be bad public policy in a health care industry which appeared to have an excess of beds. The designers of PPS, therefore, decided not to adjust payment per case for hospital size because such an adjustment might encourage inefficiency and reward construction of unneeded beds.

4. Hospital Size and Teaching Activity. When hospital size was not considered, 1981 data showed that hospital average cost per discharge increased 8.45 percent (Table ES.1) for every tenth of an intern/resident per bed; when the hospital size was considered, the increase was 5.795 percent. Congress doubled the latter figure in setting a payment adjustment of 11.6 percent; the Consolidated Omnibus Budget Reconciliation Act of 1985 reduced this adjustment to 8.1-8.7 percent.

D. THE NATURE OF OUTLIERS, SEVERITY, COMPLEXITY, AND EFFICIENCY.

1. Outliers. The DRG classification system was not designed to deal with unusual cases. The fact that there are only 468 DRGs means that uncommon cases cannot be classified into a group containing similar cases. The concept of outliers and outlier cutoffs was basic to the design of the DRGs because DRGs were designed to identify outliers, but not to explain their high costs. Congress, therefore, instructed the Department to set aside 5-6 percent of total PPS payments to pay for the "marginal cost" of "outliers" and defined outliers as patients with either unusually long stays or unusually high costs. Regulations defined length-of-stay and cost thresholds and defined marginal cost as 60 percent of the average per diem cost for nonoutlier patients for each day beyond the outlier threshold, and as 60 percent of the difference between the calculated cost and the cost threshold amount for cost outliers. There was little research basis for these decisions. Thus, the congressional mandate to examine outlier payments requires an examination of the impact of the outlier policy under actual PPS conditions and a study of the relation of outlier payments to other components of case mix adjustment.

2. Severity of Illness. Severity of illness is the patient's clinical condition at admission as indicated by the clinical likelihood of adverse outcomes such as death or disability. Severity is a somewhat ambiguous term because it includes both short-term prognosis (as in a pneumonia) and long-term prognosis (as in cancer), and because the severity of an illness for a patient includes both the illness itself and the general health of the patient. The DRG system measures severity of illness both directly, by the diagnosis recorded and by the presence of comorbidities and complications, and indirectly, through the patient's age and the specific procedures recorded. The Health Care Financing Administration (HCFA) has spent more than $10 million on severity research. The mandate to report on severity modifications to DRGs comprises determining whether suitable adjustments to DRGs are available and advisable, what impact such adjustments might have, and how valid the case mix adjustment system may be without severity adjustments if they are not feasible.

3. Intensity of Care. Intensity of care refers to the amount of medical care provided to treat patients in the same DRG who have similar severity of illness. Medically appropriate variation in intensity is extremely difficult to distinguish from inefficient and unnecessary proliferation of services. Nevertheless, the distinction is critical for payment purposes because Medicare seeks to pay for effective care and discourage inefficient and unnecessary care. This report focuses on appropriate variations in intensity and assumes that variations which are not appropriate are manifestations of inefficiency.

Intensity reflects factors such as how much diagnostic effort is necessary, whether a disease will be treated during one or several admissions, how intense an effort will be made to preserve life, and whether dying patients are treated in a hospital setting. The mandate to report on intensity adjustments to DRGs should, therefore, include an estimate of the importance of these factors, an evaluation of any available ways to quantify them, and an assessment of the adequacy of the case mix adjustment system in light of the intensity problem.

4. Efficiency. Efficiency is the relative cost of delivering a given level of medically appropriate and efficacious care to a patient. Inefficiency may result from providing excessive services, from producing individual services in an unnecessarily costly manner, or from providing services in an ineffective way. The designers of PPS believed that hospitals were not equally efficient and that incentives for efficiency were necessary to change this situation. They also realized that variations in cost that were unexplained by the PPS case measurement system might be attributable to factors other than inefficiency.

These definitions provide a framework for interpreting a series of research findings that were not available when Congress established PPS or when Medicare designed the outlier payment system. These findings result both from direct experience with hospital operations under PPS and from research conducted primarily to support this report.

II. ANALYTIC FINDINGS

A. OUTLIERS ARE A KEY TO PAYMENT EQUITY. Analysis of the distribution and impact of outlier payments shows that outliers play a critical role in the equity of PPS (Chapter 4.VI).

1. Outliers account for the majority of the difference between charges and payments under PPS. In the first year of PPS, 2.96 percent of cases were classified as outliers, but those cases accounted for 56.02 percent of the difference between

hospital charges and Medicare payments (Table ES.3). This finding suggests that more accurate outlier payments might improve the precision of PPS.

2. Outliers have very high costs, high charges, and high losses. Outlier cases are very expensive: 2.96 percent of cases account for 17.19 percent of estimated operating costs and 17.66 percent of Medicare covered charges, but only 6.35 percent of payments (Table ES.3). Average estimated costs for outliers were $16,819, which is 6.80 times the average nonoutlier cost of $2,473. The average payment for an outlier case was $7,248, resulting in an estimated loss of $9,573. This loss contrasts with an estimated profit of $791 for each nonoutlier case (Table ES.3). This finding suggests that payments for outliers could be increased without endangering the incentives to reduce costs.

3. Outliers appear to have more severe illness. Outlier cases appear to be sicker than nonoutliers, not only when their total illness is considered, but also in proportion to their longer stays (Table ES.3). Routine daily room charges are 38.0 percent of the bill for the typical patient and 21.3 percent for outliers; intensive care charges are 6.36 percent of typical bills and 11.76 percent of outlier charges; and ancillary charges are 55.7 percent of charges for nonoutliers and 66.9 percent for outliers. This finding suggests that average daily costs for stay outliers are probably more than 60 percent of average daily costs for typical cases.

4. PPS identifies stay and cost outliers in similar ways but pays more for cost outliers. Despite the very large difference in how stay and cost outliers are identified, total charges are rather similar for the two groups: $26,151 for stay

outliers and $26,012 for cost outliers, as were estimated costs ($17,116 v. $16,385). But PPS paid an estimated 51.3 percent of costs for cost outliers and 37.7 percent of costs for stay outliers (Table ES.3). This finding suggests that the distinction between cost and stay outliers may penalize hospitals when a very expensive patient stays long enough to be paid as a stay outlier.

5. PPS paid less than planned under the outlier payment system. The Department set an initial payment target of 5.7 percent of total Medicare payments. Later, however, with industry approval, the Department restricted outlier payments to the portion of payments which were not based on the hospital's historical costs, because historical costs included the hospital's tendency to have outlier patients; this lowered the target to 2.1 percent. As the 1984 Report on The Impact of the Medicare Prospective Payment System noted, actual payments under the outlier system were about 53 percent of the target. This discrepancy resulted from large decreases in average length of stay when hospitals came under PPS, which reduced the number of patients staying long enough to become stay outliers. Outlier thresholds were changed in 1985 to increase payments, but legislation delayed these changes until May 1986. The decreasing role of the hospital-specific portion in payments will also increase outlier payments per case. Available data do not show whether a change in the legislated target is appropriate.

6. Outliers are concentrated in large, urban, and teaching hospitals. Outlier cases, as well as the corresponding difference between charges and payments, were concentrated in large, urban, and teaching hospitals. This concentration of high risk cases is more than offset by the urban/rural and teaching adjustments.

Because urban/rural and teaching adjustments have been calculated without considering the benefits which outlier payments give to urban and teaching hospitals the net effect of outlier payments has been to overpay urban and teaching hospitals. The Omnibus Budget Reconciliation Act of 1986 (P.L. 99-509) changed the calculation of urban/rural payments to remove this effect, but outlier payments still overpay teaching hospitals and contribute to their high operating margins.

7. Paying "Marginal Costs" for Outliers is Impractical and Unwise. Marginal cost is the added cost resulting from services rendered after the outlier threshold is reached. The congressional mandate to pay the actual marginal cost of outlier cases presents very difficult technical problems because the date when costs are incurred cannot be determined from available data. Charge data from the first year of PPS as well as econometric studies suggest that the estimates of marginal costs used for implementation were too low, especially for cost outliers and very long stay day outliers, but precise alternative estimates are not yet possible. If PPS tried to pay exactly the full marginal costs of care beyond the outlier threshold, the inevitable imprecision of estimating actual marginal costs would mean that about half of the hospitals would actually make money on outliers once the threshold was crossed. Further, Medicare would effectively return to cost reimbursement for outlier care: there would be no incentive for hospitals to contain costs once the outlier threshold was crossed, and the outlier payment system would resemble an insurance policy for the hospital which did not require the hospital to pay coinsurance. This risk is significant because not all outlier costs result from appropriate care for which Medicare should pay. HCFA is pursuing research on ways to make payments closer to costs for patients whose costs grossly exceed the current outlier payments such as extremely expensive long stay cases.

8. Outliers may be an appropriate area in which to experiment with severity measures. Because outlier patients do have illnesses of unusual severity and because their number is small, they may be the ideal group on which first to apply measures of severity. Such measures would be particularly valuable in distinguishing between outlier cases who entered the hospital with very large identifiable needs for care and those who became outliers because of adverse events during hospitalization.

B. SEVERITY OF ILLNESS IS AN IMPORTANT DETERMINANT OF COST, BUT NO MAJOR IMPROVEMENT IN SEVERITY MEASUREMENT IS NOW AVAILABLE.

Since Congress enacted PPS, research has accumulated on the relationship of severity of illness to cost of care. Severity of illness is clearly an important determinant of cost of care in individual cases as shown in research by Brewster et al. (1985b), Coulter et al. (1985), Horn et al. (1983, 1984), the National Association of Children's Hospitals (1985), Wagner and Draper (1984), and others. The preponderance of evidence now indicates that severity is a significant determinant of differences in cost between hospitals (Cretin and Worthman 1986; Conklin et al. 1985a; Horn et al. 1985a, 1985b; Brewster unpublished data). Refinements of the DRG system appear to be feasible, but major improvements are not yet available (Chapter 3.VIII).

DRGs use discharge diagnoses, procedures performed, patient age, and (rarely) other information to assign each discharge to one of 468 groups. Although DRGs measure severity of illness through examining principal and secondary diagnosis, patient age, and some other variables, DRGs do not use physiological measures or information directly abstracted from the chart to determine severity. In contrast to some measures, the DRG algorithm and classification programs were developed with public funds and are in the public domain. HCFA has examined five systems for replacing or refining DRGs (Table ES.4):

-- Disease Staging (Staging). Staging assigns each discharge diagnosis to one of 420 staged disease categories and to one of four stages within the disease category. The stage assignment reflects the progression of the disease. As a replacement for DRGs, it has less ability to predict hospital costs accurately (Table ES.5 and Coffey and Goldfarb 1984). Replacing DRGs with Staging would discriminate against referral surgical centers because Staging would make the same payment whether or not surgery was performed. The current version of the Disease Staging classification program is proprietary. HCFA is supporting studies on using Disease Staging to refine DRGs. These studies are promising, but incomplete (Chapter 3.III).

-- Patient Management Categories (PMCs). PMCs use discharge diagnoses and procedures to assign a patient to one or more of 800 management categories. This system has been proposed as a replacement for DRGs, but has not been shown to perform better (Table ES.5). The classification program is in the public domain (Chapter 3.V).

-- Severity of Illness Index (SOII) and Computerized Severity Index (CSI). SOII is a system for assigning a severity score to a patient at discharge based on a special abstraction of the medical record. The developer has proposed using SOII within DRGs to refine DRG payment. SOII has a number of subjective elements which would create major administrative difficulties in a payment system. SOII may also tend, more than other systems, to confound care actually rendered with care needed both because it does not distinguish severity on admission from severity over the course of hospitalization and because it uses response to treatment as a measure of severity of illness. SOII has not been tested on a representative sample of hospitals, so its payment impact is not predictable. Although the SOII instrument is in the public domain, the training materials which are critical to its use are not publicly available. CSI is a newer system, developed by the same investigator, which appears to be replacing SOII. It requires special abstraction of the medical record, but the elements abstracted are combined by computer to give a score. CSI is not yet fully developed and has had only very limited tests (Chapter 3.VI). The proprietary status of CSI is not yet clear.

- Acute Physiology and Chronic Health Evaluation (APACHE). This physiologically based system assigns a patient a score based on a special abstracting of the chart. APACHE has been proposed to refine DRGs, but it has rarely been used on patients outside of intensive care units and no data are available to predict its impact on payments to hospitals. APACHE is in the public domain (Chapter 3.II).

- Medical Illness Severity Grouping System (MEDISGRPS). This system uses a special chart abstract to assign a severity score to each patient on admission to the hospital. MEDISGRPS is a new system and only preliminary data are available on its ability to predict hospital costs. MEDISGRPS is a wholly proprietary system. HCFA is supporting further studies of this system (Chapter 3.IV) at the Brandeis Health Policy Research Center.

The Department does not believe that any system to measure clinical severity of illness is currently an administratively feasible major improvement to or substitute for DRGs. Such adjustments are desirable for purposes of measuring quality — but for payment purposes this is less clear. A research strategy is in place which will result in evaluations of available severity measures over the next 3 years and may identify severity measures that can be incorporated into PPS.

We are studying DRG refinements These include:

- more sophisticated use of the additional diagnoses which are already available in the HCFA data system;
- more precise use of age in selected DRGs;
- use of source of admission for selected DRGs; and
- use of Medicare eligibility status for selected DRGs.

C. MEDICALLY APPROPRIATE VARIATIONS IN INTENSITY OF CARE APPEAR TO BE AN IMPORTANT DETERMINANT OF COSTS. A growing body of evidence suggests that medically appropriate variation in care provided to patients with comparable severity of illness accounts for significant variation in cost per case among patients, between hospital types, and across States and regions (Chapters 2.III, 4.II and 4.III). Direct methods to measure these variations are, however, much less advanced than methods to measure severity of illness and are not ready for payment applications.

1. Variation Among Patients. Patients with similar severity of illness and in similar DRGs receive care which varies in intensity for a number of medical reasons:

- Patients whose diagnosis is known before admission appear to receive fewer services than patients who must be both diagnosed and treated during the admission. Emergency admissions therefore tend to be more expensive (Munoz et al. 1985).
- Treating patients aggressively with a hope of survival or cure is usually more expensive than managing them with a goal of palliating symptoms or relieving pain. Severity of illness is only one of the influences on how

doctor and patient decide whether to take heroic measures.

-- The treatment given patients who die is much more unpredictable than the treatment of patients who live, because some die quickly with modest costs while others receive extended and heroic treatment before they die.

-- Some patients are treated in a series of admissions while others are treated definitively in a single, longer admission.

-- In general, neither the severity of illness nor the nature of the illness predicts the amount of care that will be provided to patients with chronic disease. Rather, care to these patients is influenced by what has been done in the past, what treatments remain untried, how supportive an environment the patient lives in, and what patient and doctor decide upon together.

2. Differences Across Hospital Types. Much of the difference in costs between urban and rural hospitals, teaching and nonteaching hospitals, and large and small hospitals appears to reflect differences in what is done for patients with comparable progression of illness in the same DRG (Coffey and Goldfarb 1984, Cromwell 1985)(Chapter 4.III). One basis of this conclusion is a series of studies showing that patients in less expensive hospitals have disease progression similar to those in more expensive hospitals, but that patients in less expensive hospitals are less likely to receive expensive treatments. This conclusion is analogous to studies showing that geographic variation in patterns of treatment is much greater than epidemiologic variation in underlying disease.

Other evidence points in this direction. For example, we have already noted that hospital size is significantly related to cost per discharge even after adjusting for differences in location, wage costs, and teaching intensity. The effect of size on costs largely reflects the fact that hospitals with 100-149 beds are 17.6 percent more expensive than hospitals with fewer than 51 beds (Table ES.2). By contrast, hospitals with more than 399 beds are only 6.3 percent more expensive than hospitals with 100-149 beds. The dramatically lower costs in very small hospitals are most unlikely to reflect greater efficiency; rather, these hospitals probably provide less intensive care for their patients and transfer many patients with complex illnesses to larger hospitals (Chapter 4.VII).

These findings are consistent with commonly held views that certain kinds of hospitals provide more aggressive and intensive care than other hospitals. For example, many severely ill patients are transferred to tertiary care facilities, but such transfers would appear to focus on trauma, burn, and neonatal intensive care units rather than less specialized services. Many other patients are referred to tertiary care facilities in early stages of disease either because of diagnostic uncertainty or for aggressive, technologically sophisticated or curative treatment of diseases ranging from congenital malformations to cancer. By contrast, there is no evidence that dying patients are more likely to be admitted to tertiary care facilities than to other hospitals (Chapter 4.II, 4.IV).

The availability of advanced technology, trained specialists, and experienced support staff leads to provision of medically appropriate services in some hospitals which are much less available in other hospitals. Whether large nonteaching hospitals are more like community hospitals or like large teaching hospitals remains ambiguous (Chapter 4.VII) and may depend upon the community standards to which such hospitals are held by physicians and patients.

An important result of the very different way teaching and nonteaching hospitals care for patients is that the most profitable DRGs in the former are not the most profitable in the latter (Cretin et al. 1986). Knowing the relative profitability of a DRG in a teaching hospital predicts less than half of the variation in its relative profitability in a nonteaching facility. This suggests that teaching hospitals are not simply doing more, but practicing a different pattern of medicine. Available data do not permit classifying more intense care as either having higher quality or being less efficient.

3. Differences Across States and Regions. Health services appear to be organized and delivered in ways which vary dramatically across States and regions, and these differences appear to affect costs (Chapter 4.III):

-- The number of beneficiaries who die within 30 days of admission does not vary greatly across States, but the relative number of beneficiaries who die inside or outside of a hospital during that period depends heavily on average length of hospital stay in the State. In States where average stays are short, more patients appear to receive terminal care in nonhospital settings (Table ES.6).

-- The rate at which surgery is performed varies widely across small and large regions. Likewise, the rate of hospital admissions varies across regions. Areas in which hospital admission rates are lower tend to have higher costs per admission (Gornick 1982). Although severity of illness may account for a portion of this variation, there is no evidence that it does so. Likewise, there is as yet no evidence that appropriate criteria for admission can be made sufficiently clear to abolish these forms of interstate variation, even though they can almost certainly be reduced.

D. PPS PAYS APPROPRIATELY FOR CARE TO VULNERABLE BENEFICIARY GROUPS. PPS does not appear to create payment incentives which discriminate against vulnerable beneficiary groups such as the poor, the very old, or the disabled (Chapter 3.VII). Charges for admissions of beneficiaries over 80, over 85, previously hospitalized, and disabled and charges for admissions of beneficiaries also eligible for Medicaid are very similar to charges for admissions of other Medicare beneficiaries to the same hospital and in the same DRG (Table ES.7). The greatest discrepancy is that charges for patients admitted from nursing homes are about 7 percent higher than charges for other beneficiaries. Despite the generally small differences found, a number of these patient characteristics might be incorporated into the DRG system to improve specific DRGs. The findings are restricted to differences between patients in the same hospital and do not provide information on the relative costs of hospitals which have a disproportionate share of any of these beneficiary groups.

E. EVEN WITHOUT A MORE REFINED CASE MIX ADJUSTMENT, HOSPITALS GENERALLY DID WELL IN THE FIRST YEAR OF PPS. Although the findings summarized to this point raise the possibility that unmeasured severity and intensity and low outlier payments may compromise the PPS case mix adjustment system, empirical data are necessary to determine whether the various other adjustments in PPS compensate for these problems.

HCFA examined Medicare cost reports for hospitals' first year under PPS and the impact of recent payment changes on the payment which would be made under fully national rates starting in 1988. Some of the cost reports were unaudited, but analysis suggested that the overall picture was similar to that which could be expected from the final audited reports.

1. Operating Margins. In the first year of PPS, Medicare payments for operating expenses of inpatient care exceeded operating costs for that care by 16.4 percent. Under national rates, we project that Medicare payments for inpatient care will exceed hospital costs by 12.4 percent. Because smaller hospitals tended to have lower margins than larger hospitals, the average hospital had a Medicare

operating margin of 12.01 percent in its first PPS year and has a projected operating margin of 12.14 percent under national rates. The American Hospital Association reports sharply lower overall hospital operating margins, but the differences appear to reflect different accounting procedures and the fact that HCFA data is limited to Medicare patients'. Hospitals generally realize greater margins on Medicare beneficiaries than on other patients.

Projections for future years are based on an assumption that the only change in cost per case will be that resulting from inflation in the hospital market basket. However, hospitals can and do change their costs dramatically in response to PPS incentives. In the first year of PPS, the average hospital reduced its costs 16.4 percent compared to the costs which HCFA had anticipated (because HCFA overestimated costs, the actual economies were somewhat less). Under fully national rates, and assuming that operating costs increase exactly at the rate of inflation, we project that 23.3 percent of hospitals will have an operating loss, while 26.0 percent will have an operating profit greater than 30 percent.

The profitability of Medicare business resulted from four factors which may not be repeated:

- The hospital marketbasket inflated less than expected.
- Changes in diagnosis and procedure coding increased DRG-based payments more than had been expected.
- Pressure of time forced HCFA to set payments by using unaudited cost reports which overstated base year costs.
- Hospitals were able to achieve significant efficiencies under the new PPS incentives.

To bring Medicare payments more in line with costs, the Department has proposed PPS update factors to correct for overestimate of inflation, changes in coding, and elimination of inefficient practices. Congress has enacted higher updates than the Department has proposed.

2. Overall Correlation Between Costs and Payments. In the first year of PPS, payments were closely correlated with costs. Average payment per case at the hospital level reflected 89.0 percent of the variation in hospital average cost per case. Under fully national rates, the correlation is less close, but we project that payments per case at the hospital level will still reflect 71.4 percent of variation in hospital average cost per case.

3. The Range of Variation in Operating Margin Among Hospitals. Among hospitals there are wide variations in operating margins. In the first year of PPS, 19.4 percent of hospitals had an operating loss while 9.5 percent had an operating profit greater than 30 percent. Under fully national rates, we project that 23.3 percent of hospitals will have an operating loss, while 26.0 percent will have an operating profit greater than 30 percent. In the following analysis we discuss the variation in operating margin across classes of hospitals, but the reader should notice that Tables ES.1 and ES.2 show that the relation of hospital characteristics to costs (and hence to margins) depends heavily on which combination of characteristics which we consider.

4. Urban/Rural. In the first year of PPS, Medicare losses were concentrated in rural hospitals, but the Omnibus Budget Reconciliation Act of 1986 made changes in the payment formula which eliminated this difference. Under fully national rates, we project that 23.2 percent of rural hospitals (compared to an overall rate of 23.3 percent) will have an operating loss, and 26.2 percent of rural hospitals (compared to an overall rate of 26.0 percent) are projected to have an operating profit greater than 30 percent.

5. Teaching/Nonteaching. In the first year of PPS, teaching hospitals did substantially better than nonteaching. While 22.4 percent of nonteaching hospitals had losses and 7.9 percent had profits above 30 percent, only 5.4 percent of light

teaching hospitals (less than 0.25 residents per bed) and 4.10 percent of heavy teaching hospitals has losses, while 14.4 percent of light teaching hospitals and 32.0 percent of heavy teaching hospitals had profits above 30 percent. Recent legislation has reduced the teaching adjustment, adjusted the outlier system, changed the urban/rural differential, and created a disproportionate share adjustment, resulting in a net reduction in the advantages of teaching hospitals. We project that, under fully national rates, 23.4 percent of nonteaching hospitals, 23.8 percent of light teaching hospitals, and 18.7 percent of heavy teaching hospitals will have operating losses; 26.7 percent of nonteaching hospitals, 20.4 percent of light teaching hospitals, and 33.7 percent of heavy teaching hospitals will have operating profits above 30 percent.

6. Hospital Size. Hospital size is directly related to profitability, but the relationship reverses between the first year of PPS and projected experience under national rates. In the first year, for example, 35.1 percent of hospitals with fewer than 50 beds had an operating loss, while 8.6 percent had an operating profit over 30 percent; no hospital over 599 beds had a loss in the first year while 26.7 percent had an operating profit over 30 percent. By contrast, we project that under national rates 21.1 percent of hospitals with fewer than 50 beds would have an operating loss while 39.3 percent would have an operating profit above 30 percent; 18.1 percent of hospitals over 599 beds would have a loss while 21.6 percent would have more than 30 percent profit. The reversal appears to reflect the lower historical costs of small hospitals and their inability to cut costs as much as larger hospitals. In the first year of PPS, high operating margins resulted mainly from hospitals doing well compared to their own historical costs; at full implementation, high operating margins will reflect how hospitals do compared to other hospitals.

7. Variation Across Regions. Operating margins vary across regions and will

vary more under fully national rates, even after adjusting for differences in hospital size, teaching activity, case mix, and wages. Table ES. 10 shows that in the first PPS year the mid-Atlantic States, Illinois, and Michigan had operating margins about 4-5 percentage points higher than New England, Kentucky, Tennessee, Alabama, Alaska, and Hawaii, with margins in other States falling between. We project that under fully national rates hospitals in Iowa, Kansas, Mississippi, Missouri, Nebraska, North Dakota, South Dakota, and Utah will have margins about 10 percentage points higher than comparable hospitals in Arizona, Florida, Illinois, Michigan, Nevada, and New Mexico. Because these differences reflect differences in margins for hospitals with similar characteristics while average hospitals have different characteristics in different States, the table does not portray differences between the average hospital in different States. The difference in patterns from the first PPS year to national rates reflects the same effects noted in discussing large and small hospitals: hospitals which were historically expensive do well when they compete against their record but much less well than they compete against national standards.

F. PPS DATA INTEGRITY IS GOOD. The data on which the PPS case mix adjustment system operates appear to have enough integrity for the equitable operation of the program (Chapters 2 and 3). The substantial change in coding which occurred at the time of implementation of PPS is consistent with improved data quality, (Chapter 2.IV.F). DRG weights may be too low for complex cases and too high for simple cases, but it is not clear that this "weight compression" has actually resulted in inaccurate payments (Chapter 4.IV). Although the data used to calculate the original DRG weights included significant amounts of unnecessary care, that data weakness does not appear to have resulted in inaccurate DRG weights (Chapter 4.IV).

G. RESEARCH EFFORT. Before Congress enacted PPS, HCFA spent more than $9.4 million on case mix research; HCFA has spent $5.1 million more since then. Although this report documents impressive progress on a number of fronts, there is no one "solution" to the problem and the need for a number of partial solutions is increasingly clear. HCFA has already invested heavily in a number of the most promising options.

H. QUALITY OF CARE AND CASE MIX ADJUSTMENT. Case mix adjustment has both a direct and an indirect application to quality of care. Directly, more accurate case mix adjustment will permit paying hospitals the amounts necessary to deliver quality care and to assure beneficiaries access to quality care. Indirectly, and perhaps equally important, the same measures of case mix which permit more accurate payment to hospitals also permit more accurate comparisons of health care outcomes across hospitals by allowing investigators to assure that comparisons between hospitals are made for patients who are truly similar.

I. CASE MIX ADJUSTMENT AND CAPITATION RATES. Research on case mix adjustment promises to make several contributions to setting more accurate payments for capitated health care systems such as HMOs. While case mix refinement based on acute severity of illness has limited application to capitated rates, study of the way chronic illness affects hospital costs offers direct application to setting capitation rates.

III. CONCLUSIONS

The findings lead to three significant policy conclusions:

A. THE MEANING OF DIFFERENCES BETWEEN COSTS AND PAYMENTS. Because we do not have satisfactory measures of severity of illness or satisfactory ways to assess what intensity of care is medically appropriate, we cannot resolve the questions of which differences between costs and payments are appropriate. Clearly, perfect correlation between costs and payments would be equivalent to cost-based reimbursement; because there still appears to be substantial room for increasing efficiency in the hospital industry, payments which perfectly reflect costs would not be a desirable goal. On the other hand, we cannot reach a technical judgment whether the degree of match between costs and payments which we have found, and which we project to occur under national rates, is the correct one. The differences between costs and payments may reflect correctable inefficiencies for which Medicare should not pay; if this is true, then the high operating margins realized by some hospitals may be useful incentives to other hospitals to become more efficient. On the other hand, the differences between costs and payments may mean that hospitals with severely ill patients and medically appropriate intense care are being penalized by a system which does not fully measure their special responsibilities. Current technology does not allow us to determine which of these two scenarios is most correct.

During the first years of PPS, most hospitals were protected from possible payment imprecision under national rates not only by the blend of regional and hospital-specific rates but also by high payments relative to historical costs and substantial gains in efficiency. In the immediate future, a growing number of hospitals may be pressed by an increasingly competitive health care environment. Hospitals will have fewer alternative sources of income to which they can turn if Medicare's payments are too low. On the other hand, Medicare currently provides average margins in excess of 12 percent, which leaves substantial protection against any error.

Under these circumstances, continuing close monitoring of the impact of PPS on providers and patients becomes especially important, and close cooperation between Congress and the Administration in charting the future course of the program will be essential. The refinements of case mix measurement which could improve payment for severity and complexity are also the research tools necessary to determine whether PPS payments are equitable; further refinement of case mix adjustment methods is an integral part of the future of PPS.

B. IMPROVING CASE MIX ADJUSTMENT THROUGH OUTLIER POLICY. One way to improve accuracy of payment at the individual case level is to pay more precisely for outlier cases. Current law, which requires paying the full marginal costs of care of outliers, raises technical problems which cannot be fully resolved and provides no incentives for efficiency once a case passes the outlier threshold. HCFA is actively pursuing research on refining outlier payments.

C. IMPROVING CASE MIX ADJUSTMENT THROUGH DRG REFINEMENT. Although major immediate improvements in the DRG system are not possible, modest improvements have already been made and others appear feasible. The latter include using Disease Staging to refine DRGs, more sophisticated use of age in DRGs, and better use of comorbidities and complications. In the longer term, use of physiologic measures and further adjustment for intensity of care appear to be productive strategies.

IV. FURTHER PLANS

These findings and conclusions lead to the following program:

A. CONTINUE THE PRESENT DRG UPDATE SYSTEM. HCFA will continue its present system for DRG updates, which has led to appropriate progressive changes in the system.

B. HCFA will continue to study:

- Approaches for refinements to DRGs based on more precise use of comorbidity and complication data. These refinements will develop from work at Systemetrics using Disease Staging, work at Yale using a variety of other systems, or some combination of the two.
- Approaches for changes in teaching adjustments to take account of outlier payments.

C. SUBSEQUENT REFINEMENTS:

- Evaluate the application of severity of illness measures to payment for outlier cases.
- Analyze pilot studies of severity measures to determine whether and at what cost they provide a general improvement in case mix adjustment.

V. ORGANIZATION OF THE REPORT

This report is designed to develop a framework for analyzing outliers, severity, and intensity; to present a substantial body of new research, much of it carried out for this report; and to document the detailed evidence and reasoning behind the recommendations contained in this Executive Summary. The report provides:

- a review of the goals and organization of PPS (Chapter 1);
- a discussion of the reasons for variation in need for care among individual patients and the problems of measuring that need (Chapter 2);
- a summary of all available research on alternatives to and improvements to DRGs (Chapter 3);
- an examination of the reasons for variation in average need for care among the patients of different hospitals and of the impact of PPS and outlier payments on hospitals (Chapter 4); and
- an agenda for improving case mix measurement in the next 5 years (Chapter 5).

Table ES.1

PREDICTORS OF HOSPITAL AVERAGE COST PER MEDICARE CASE
Percentage Increase for Every 1.0 Percent Increase in Variable

	1981 Costs	1984 Costs	
A: Basic Model			
Case Mix Index	1.014	0.954	
Teaching*	0.579	0.583	
Wage Index	1.019	0.959	
Large SMSA	11.60	9.91	)percent
Medium SMSA	2.76	3.78	)higher
Small SMSA	0.10	2.54	)than rural
Number of Beds	0.119	0.086	
R^2	0.725	0.702	
B: Without Hospital Size (beds)			
Case Mix Index	1.464	1.295	
Teaching*	0.845	0.766	
Wage Index	0.980	0.863	
Large SMSA	23.18	19.97	)percent
Medium SMSA	11.96	11.34	)higher
Small SMSA	9.07	9.35	)than rural
R^2	0.692	0.682	
C: Coefficients Reflecting Actual 1984 Payment Procedures			
Case Mix Index	1.000[1]	1.000[1]	
Teaching*	1.158[1]	1.158[1]	
Wage Index	0.7915[1]	0.7915[1]	
Urban	25.34	18.70	percent higher than rural
R^2	0.666	0.669	

[1] Value mandated by law or regulation.

* Teaching is defined as the ratio of interns and residents to hospital beds.

1981 (N=4971): hospitals with >49 Medicare discharges in 1981 and any 1984 discharges.

1984 (N=5349): hospitals with >49 Medicare discharges and with available Cost Reports.

(continued)

Table ES.1
(continued)

PREDICTORS OF HOSPITAL AVERAGE COST PER CASE
Percentage increase for Every 1.0 Percent Increase in Variable

	1981 Costs	1984 Costs
D: With Hospital Bed Size Groups		
Case Mix Index	1.097	1.003
Teaching*	0.742	0.702
Wage Index	1.001	0.943
Large SMSA	13.21	10.57)) percent
Medium SMSA	3.91	4.42)) higher
Small SMSA	1.04	2.87)) than rural
51-100 Beds	14.11	9.10) percent
101-250 Beds	25.24	18.27) higher than
251-500 Beds	29.55	21.07) 1-50 bed
More than 500 beds	32.81	21.15) hospitals
R^2	0.725	0.702
E:		
Case Mix Index	2.851	2.298
R^2	0.401	0.423
F:		
Case Mix Index	1.469	1.376
Teaching*	0.899	0.846
Wage Index	1.436	1.191
R^2	0.670	0.669

* Teaching is defined as the ratio of interns and residents to hospital beds.

1981 (N=4971): hospitals with >49 Medicare discharges in 1981 and any 1984 discharges.

1984 (N=5349): hospitals with >49 Medicare discharges and with available Cost Reports.

Source: Medicare Cost Reports.

Table ES.2

PREDICTORS OF HOSPITAL AVERAGE COST PER MEDICARE CASE IN 1984

Variable	Effect[1]	Cumulative[2] R^2
Wage Index	0.971	0.487
Case Mix Index	1.275	0.641
Teaching Activity	0.686	0.661
Urban location	11.068	0.672
Disproportionate share[3]	0.719	0.674
Including Outlier Payments		
Wage Index	0.891	0.487
Case Mix Index	1.253	0.641
Teaching Activity	0.336	0.661
Urban location	6.716	0.672
Disproportionate share[3]	0.133	0.674
Outlier payments[4]	2.571	0.737
Controlling for Hospital Size		
Wage Index	1.031	0.482
Case Mix Index	0.951	0.641
Teaching Activity	0.618	0.661
Urban location	3.821	0.672
Disproportionate share[3]	0.497	0.674
50-99 beds	9.771	0.696
100-149 beds	17.626	
150-199 beds	18.338	
200-399 beds	22.308	
more than 399 beds	25.028	

[1] Effects correspond to coefficients in a log-linear regression model which includes all of the variables. For continuous variables (wage index, case mix index, teaching activity, disproportionate share, and outlier payments) the effect is the percent increase in cost corresponding to a 1.0 percent increase in the variable. For dummy variables (urban location and bed size) the effect is the percent increase in cost per case if hospital is urban or if it has the indicated number of beds, compared to a rural hospital or to one with 1-49 beds.

[2] Cumulative R^2 is the portion of the total variance in hospital average cost per case which is accounted for in a stepwise regression by variables down to the given value. All hospital size dummies are entered together.

[3] The disproportionate share variable is the percentage increase in payments to the hospital which is defined by legislation; the coefficient indicates the percent increase in cost which corresponds to a 1.0 percent increase in payments required by law.

[4] Outlier payments are the fraction of payments for the hospital made to the hospital estimated from casemix during the first year of PPS and payment rules for Federal Fiscal Year 1986 but assuming full implementation of PPS.

Source: Medicare Cost Reports. N=5350 hospitals.

Table ES.3

Characteristics of Medicare Outlier Cases
Cases Paid Under PPS in Federal Fiscal Year 1984

	Type of Outlier				
Variable	Length of Stay	Cost	All	Non-Outliers	All PPS Cases
PPS Discharges	1.76%	1.21%	2.96%	97.04%	100.00%
Urban Hospitals	2.07	1.45	3.52	96.48	100.00
Rural Hospitals	0.98	0.60	1.58	98.42	100.00
Teaching Hospitals	2.58	1.24	3.82	96.18	100.00
Nonteaching Hospitals	1.34	1.19	2.53	97.47	100.00
Hospital Days	9.17	3.34	12.51	87.49	100.00
Total Charges	10.49	7.17	17.66	82.36	100.00
Est. Operating Costs	10.38	6.80	17.19	82.81	100.00
Medicare Payments	3.35	3.00	6.35	93.65	100.00
Charges less Payments	34.71%	21.32	56.02	43.98	100.00
Special care charges (percent of total)	10.10	14.20	11.76	6.36	7.32
Ancillary charges (percent of total)	62.12	63.89	62.83	55.68	56.48
Length of Stay	40.9	21.7	33.1	7.1	7.8
Total Charge	$ 26,151	$26,012	$26,094	$ 3,714	$ 4,377
Charge per Day	$ 639	$1,200	$788	$ 526	$ 559
Est. Operating Cost	$ 17,116	$16,385	$16,819	$2,473	$2,898
Medicare payment	$ 6,457	$8,399	$7,248	$3,262	$3,380
Est. Operating Margin	$-10,658	$-7,989	$-9,573	$+791	$+484

Source: Stuart Guterman, Health Care Financing Administration, unpublished data.

Table ES.4

BASIC FEATURES OF ALTERNATIVE CASE MIX MEASUREMENT SYSTEMS

SYSTEM NAME	CREATOR	ORGANIZATION	ORDINAL/ NON-ORDINAL*	NUMBER OF GROUPS/SCALE POINTS	DATA REQUIREMENTS	PROPRIETARY
Diagnosis-Related Groups (DRGs)	Thompson Fetter Averill	Yale	Nonordinal	468	UHDDS	No
Acute Physiology and Chronic Health Evaluation (APACHE)	Knaus Wagner	George Washington University	Ordinal	71	specially abstracted physiologic data	No
Disease Staging (Staging)	Gonella	Thomas Jefferson University	Nonordinal**	420 groups 4 levels each)	UHDDS	Software yes; architecture no
Medical Illness Severity Grouping System (MEDISGRPS)	Brewster	MediQual Inc.	Ordinal	5	specially abstracted history and physiologic data	Software and architecture secret.
Patient Management Categories (PMCs)	Young	Blue Cross of Western Pennsylvania	Nonordinal***	720	UHDDS	No
Severity of Illness Index (SOII)	Horn	Johns Hopkins University	Ordinal	5	specially coded scales "implicitly" integrated	No, but training manuals not public
Computerized Severity Index (CSI)	Horn	Johns Hopkins & Health Systems International	Nonordinal**	5	specially abstracted physiologic data	Incomplete

*In an ordinal system, patients have greater severity or care needs than patients with lower numbers and less than patients with higher numbers.
** Stages are ordinal within each staged disease category. CSI levels are ordinal within a diagnosis.
*** PMCs are ordinal within major disease category.

Table ES.5

COMPARISON OF THE VARIATION (R-SQUARE) IN COST OF INDIVIDUAL CASES EXPLAINED BY THREE CASE MIX MEASURES

	Diagnosis-Related Groups	Patient Management Categories	Disease Staging
Trimmed	0.30	0.26	0.17
Untrimmed	0.17	0.15	0.10

* Cost data are untrimmed and exclude DRGs 468 and 470.

Source: Calore, unpublished data 1985.

Table ES.6

DEATHS DURING ADMISSION AND WITHIN 30 DAYS AFTER DISCHARGE
5 Percent Sample of 1984 Medicare Discharges

	Total Adms	Deaths Within 30 Days	Deaths During Admission	Difference	Average Length of Stay
Oregon	5995	8.4%	4.9%	3.5%	6.51
Utah	1897	8.8%	5.6%	3.2%	6.64
Wyoming	725	7.9%	4.8%	3.0%	7.52
Washington	8292	8.0%	5.3%	2.7%	6.86
Montana	1802	8.2%	5.7%	2.6%	7.00
Iowa	7573	7.4%	5.0%	2.4%	8.12
Kansas	6466	7.5%	5.1%	2.4%	8.67
Minnesota	9503	7.6%	5.2%	2.4%	7.69
Nevada	1691	8.6%	6.3%	2.4%	7.81
Arkansas	7385	8.3%	5.9%	2.4%	7.30
S Dakota	2214	6.0%	3.7%	2.2%	7.45
Wisconsin	11278	7.4%	5.2%	2.2%	8.35
Arizona	5946	7.7%	5.4%	2.2%	7.92
Oklahoma	7779	8.5%	6.3%	2.2%	7.75
N Mexico	2246	6.9%	4.7%	2.2%	7.32
Colorado	4692	7.1%	4.9%	2.2%	7.84
Nebraska	4491	7.7%	5.6%	2.1%	7.94
Georgia	13633	7.4%	5.4%	2.0%	7.60
Missouri	15294	7.6%	5.5%	2.0%	8.94
Kentucky	9256	7.5%	5.5%	2.0%	8.46
Texas	33071	7.5%	5.5%	2.0%	8.14
N Dakota	2260	6.3%	4.4%	1.9%	7.79
California	44630	7.6%	5.7%	1.9%	7.71
Louisianna	9814	7.4%	5.5%	1.9%	7.85
Idaho	1807	6.8%	4.9%	1.8%	6.50
Tennessee	14757	7.0%	5.2%	1.8%	8.65
Indiana	12871	7.6%	5.8%	1.8%	8.89
Vermont	784	7.9%	6.1%	1.8%	8.93
Alablam	10603	6.6%	4.8%	1.8%	8.15
Mississippi	7741	6.8%	5.1%	1.8%	7.76
Ohio	21572	7.6%	5.9%	1.8%	9.32
Delaware	1201	8.7%	7.0%	1.7%	10.41
Florida	34209	7.3%	5.6%	1.7%	8.42
Hawaii	1156	8.0%	6.3%	1.6%	8.74
N Carolina	12150	8.2%	6.6%	1.6%	8.83
Illinois	27084	7.5%	5.9%	1.6%	9.72
S Carolina	6282	7.2%	5.8%	1.5%	8.84
Penna.	34868	7.3%	5.9%	1.4%	10.03
Michigan	19938	7.3%	5.9%	1.4%	9.31
New Hamp.	1363	6.9%	5.5%	1.4%	8.61
Virginia	10678	7.5%	6.1%	1.3%	9.28
Maine	3076	7.3%	6.0%	1.3%	8.44

(continued on next page)

Table ES.6
(Continued)

DEATHS DURING ADMISSION AND WITHIN 30 DAYS AFTER DISCHARGE
5 Percent Sample of 1984 Medicare Discharges

	TOTAL ADMS	DEATHS WITHIN 30 DAYS	DEATHS DURING ADMISSION	DIFFERENCE	AVERAGE LENGTH OF STAY
W. Virginia	5769	6.7%	5.4%	1.2%	8.17
Rhode Is.	2193	7.7%	6.6%	1.1%	10.33
Connecticut	6392	8.1%	7.1%	1.0%	9.84
Maryland	6107	7.4%	6.7%	0.8%	10.33
Dist. Col.	1293	6.7%	6.0%	0.7%	12.20
Mass.	14560	7.2%	6.7%	0.5%	11.58
Alaska	226	4.0%	3.5%	0.4%	7.96
New Jersey	15837	7.7%	7.6%	0.1%	12.11
New York	37973	7.8%	8.6%	-0.8%	13.82
U.S.	530423	7.5%	5.9%	1.6%	9.11

Source: Eggers P. Unpublished Data.

Table ES.7

RELATIVE CHARGES FOR BENEFICIARIES WITH EIGHT CHARACTERISTICS
Compared to Otherwise Similar Beneficiaries
Controlling for DRG and Hospital in which Care was Rendered
5 Percent Sample of 1983 Medicare Discharges

Beneficiary Characteristic	Relative Charge
1. Nursing Home Resident.	1.062***
2. Hospitalization in Previous Year.	1.033***
3. Medicaid Enrollee	0.985**
4. Aged (>79).	1.038***
5. Very aged (>84).	0.987*
6. Younger disabled <65)	0.993
7. Younger endstage renal disease.	1.052**
8. Older disabled. (>64)	1.067***

* = $p <.05$ that ratio is = 1.00; ** = $p <.01$ that ratio is = 1.00;
*** = $p <.001$ that ratio is = 1.00, using t test.

[1] Based on 10 random samples of 60 DRGs each.

Source: Jencks & Kay 1986.

Table ES.8

Medicare Operating Margin By Urban/Rural Location and Number of Beds

Actual First Year of PPS and Projected Experience under Full National Rates

		Percent of Hospitals									Number of Hosps
		Loss				Profit					
		>20%	20-10%	10-5%	5-0%	0-5%	5-10%	10-20%	20-30%	>30%	
Urban <100 beds	PPS yr 1	4	4	5	6	11	11	28	19	13	545
	National rates	6	5	5	6	8	7	14	10	40	675
Urban 100-404 beds	PPS yr 1	1	1	1	4	8	13	35	25	12	1347
	National rates	3	6	5	9	10	10	20	14	21	1649
Urban 405-684 beds	PPS yr 1	0	0	0	3	5	13	32	29	18	277
	National rates	0	8	5	8	11	11	21	17	20	332
Urban >684 beds	PPS yr 1	0	0	0	0	1	3	15	16	13	48
	National rates	0	10	7	3	11	11	21	17	20	70
Rural <100 beds	PPS yr 1	6	8	7	11	12	12	26	12	7	1795
	National rates	4	7	5	6	8	9	18	14	30	2029
Rural 100-169 beds	PPS yr 1	1	4	7	10	17	17	31	11	3	357
	National	4	10	6	11	11	8	20	15	13	397
Rural >169 beds	PPS yr 1	1	3	4	10	16	22	31	10	3	175
	National	4	7	5	7	11	10	26	14	17	197
All hosps	PPS yr 1	3.0	4.4	4.5	7.6	10.8	12.6	29.9	17.7	9.5	4544
	National	3.7	6.9	5.2	7.5	9.0	9.4	18.6	13.8	26.0	5349

Table ES.9

Medicare Operating Margins By Urban/Rural Location and Teaching Status

Actual First Year of PPS and Projected Experience under Full National Rates

		Percent of Hospitals									Number of Hosps
		Loss				Profit					
		>20%	20-10%	10-5%	5-0%	0-5%	5-10%	10-20%	20-30%	>30%	
Urban non-teaching	PPS yr 1	2	2	3	5	10	13	32	22	11	1491
	National	4	6	6	8	10	9	18	13	27	1823
Urban light teaching	PPS yr 1	0	1	1	3	6	12	27	37	15	608
	National	2	7	6	9	9	12	20	13	21	744
Urban heavy teaching	PPS yr 1	0	1	0	3	4	3	27	30	31	118
	National	2	6	5	6	8	7	21	12	33	159
Rural non-teaching	PPS yr 1	5	7	7	11	13	13	27	12	6	2271
	National	4	7	5	7	9	9	19	14	26	2556
Rural light teaching	PPS yr 1	2	4	2	2	17	15	37	13	8	52
	National	3	10	8	2	8	5	27	20	17	60
Rural heavy teaching	PPS yr 1	0	0	0	0	0	25	0	25	50	4
	National	0	14	0	0	0	14	0	14	57	7
All hosps	PPS yr 1	3.0	4.4	4.5	7.6	10.8	12.6	29.9	17.7	9.5	4544
	National	3.7	6.9	5.2	7.5	9.0	9.4	18.6	13.8	26.0	5349

Table ES.10

Variations in Hospital Medicare Operating Margins Across Regions Controlling for other hospital characteristics

Variable	PPS Year 1[1]	National Rates[1]
New England	0.00	0.00
Mid Atlantic	+5.43****	+7.71****
South Atlantic ex FL	+1.02	+9.55****
South Atlantic: FL	-0.10	-0.59
East N Cent: IL & MI	-4.89****	-1.41
East N Cent: IN, OH, WI	3.63***	6.56****
East S Cent: MS	3.43*	14.48****
East S Cent: KY, TN, AL	-0.16	10.80****
West N Cent: NB, ND, SD	2.38	16.36****
West N Cent: IO, KA, MO, MN	1.47	9.07****
West S Cent: TX	1.08	7.71****
West S Cent: AR, LA, OK	1.09	11.11****
Mountain: UT	2.42	14.53****
Mountain: AZ, NM, NV	-1.81	-1.03
Mountain: ID, MT, CO, WY	3.52**	10.62****
Pacific: CA	1.77	3.93*
Pacific: AK, HI	-2.81	7.37
Pacific: OR, WA	3.14*	16.35****
Wage Index	-0.084**	-0.147***
Case Mix Index	0.382****	-0.011
Teaching Intensity	0.097**	0.182***
Urban	6.097****	9.359****
Disproportionate share[2]	0.173*	0.479****
Less than 50 beds	0.00	0.00
50-99 Beds	2.498****	-5.887****
100-149 beds	1.113	-10.841****
150-199 beds	0.476	-9.134****
200-399 beds	0.501	-12.012****
More than 399 beds	0.728	-13.669****
Adjusted R^2	0.163	0.091

*= p <.05; ** = p<.01; *** = p<.001; **** = p<.0001

For notes see Table ES.2. Coefficients for regions are relative increase or decrease in operating margin compared to a otherwise similar hospital in the New England region.

Source: Medicare Cost Reports.

CHAPTER 1

THE GOALS AND METHODS OF THE PPS CASE MIX ADJUSTMENT SYSTEM

Case mix adjustment is the procedure by which Medicare adjusts per case payments to an individual hospital to reflect the characteristics of cases treated in that hospital. Case mix adjustments may also be made to compare hospital mortality rates and other measures of hospital performance, but this report discusses such potentially important uses only briefly.

There is no "gold standard" against which to judge the adequacy of a case mix adjustment system, and neither legislation nor regulation specifies the exact purposes of case mix adjustment under PPS. To decide whether case mix adjustment is doing what it should, we must, therefore examine many different kinds of evidence and fit them together into a complete picture. This chapter reviews case mix adjustment under PPS, and identifies the goals the system should achieve and specific criteria against which it should be judged.

Case mix measurement and the payment system are inseparable. DRGs can be evaluated and refined only as part of a system which includes outlier payments and hospital-level payment adjustments. The DRG classification system was never designed or expected to reflect payments for outliers or to take account of factors such as wages, which influence local costs. The PPS outlier payment was designed to deal with inherent limitations of DRGs, and hospital-level adjustments were designed to complement the DRG and outlier "systems". This interdependence dictates the structure of this chapter. Section I describes the goals of case mix adjustment, Section II examines the interdependent system which seeks to meet these goals, and Section III discusses the significance of differences between a hospital's costs and what Medicare pays and describes ways in which different parts of the payment system can interact to produce complex payment effects.

I. THE GOALS OF CASE MIX ADJUSTMENT

PPS created a revolution by paying hospitals for the care that patients were expected to need on average rather than the care that the hospital actually provided. This involved two separate reforms--paying for the case rather than for individual services (paying for larger aggregations of services is often called bundling) and defining the payment in a schedule published before care is rendered (prospective payment).

A. GOALS OF PPS. PPS has three primary goals:

1. Promote Efficiency. Give hospitals incentives to provide quality care as efficiently as possible by eliminating per-service retrospective payments, which encouraged long stays and profligate use of services.

2. Control Costs. Make Medicare's payments for hospital care predictable and controllable through prospectively determined payments for a unit of service (the admission) whose volume and composition has historically been relatively predictable.

3. Be Equitable. Make resource allocation equitable to hospitals and patients by basing allocation on actual patient need rather than on possibly inappropriate historical patterns.

There is no serious question that PPS has succeeded in its first two goals. However, this success does not necessarily imply equity; nor does initial success necessarily imply long-term viability. During the phase-in of PPS, hospitals have been protected against sharp changes in Medicare payments by a hospital-specific component and a regional component in their payment rates. In addition, a teaching adjustment and an urban/rural adjustment were included to further reduce sudden changes in payment. Hospitals were further protected by three "one-time" events which resulted in payments which were generally higher than costs:

- -- the initial costs on which payments were based were overestimated because audited cost reports were unavailable,
- -- improved coding accuracy significantly increased payments, and
- -- hospitals were able to realize substantial operating efficiencies.

If as overpayments are corrected and hospital-specific and regional payments are phased out, hospitals will have less protection against possible imprecision in the system. At the same time, hospitals will face private sector competition, problems with uncompensated care, and a Federal budget constrained by "Gramm-Rudman-Hollings."

In the short run, PPS creates strong incentives for efficiency independent of its equity. A hospital has powerful reasons to shorten inpatient stays, to provide some services for the inpatient stay prior to admission or after discharge, and to provide a more efficient mix of services because it knows that each dollar spent on care reduces its profits. PPS encourages efficiency even if it does not pay hospitals exactly what they sometimes need to treat their patients, but chronic underpayment could decrease access to care and quality of care.

B. CRITICISMS OF CASE MIX ADJUSTMENT. When Congress asked for a report on severity of illness, intensity of care, and outlier payments, the clear concern was whether the PPS case mix adjustment system would pay fairly for the costs of sophisticated, complex, and critical care. Hospitals, physicians, and others have charged that the current case mix measurement system fails in three important ways to measure patients' needs for care:

1. Severity. Critics charge that the system fails fully to measure severity of illness and need for costly specialized and tertiary care, resulting in systematic underpayments to hospitals specializing in such care. Specifically, they argue that

the higher costs of teaching hospitals and the lower costs of small hospitals reflect differences in average severity of patient illness which should be measured by the case mix adjustment system.

2. Special Populations and Diseases. Critics charge that special costs of caring for the frail, the very old, the disabled, patients with special diseases, and the poor are not fully measured, giving hospitals incentives to discriminate against such patients. As specific evidence, critics cite findings that the costs of hospitals with a disproportionate share of poor patients are higher than the case mix system predicts. These higher costs led Congress to legislate in 1986 adjustments for disproportionate share hospitals similar to those enacted for teaching and urban/rural differences (Consolidated Budget Reconciliation Act of 1985 (COBRA), P.L 99-272).

3. Random and Systematic Risk from Severely Ill or "Outlier" Patients. Critics charge that Medicare makes the same basic payment for patients who vary so much that hospitals, especially small hospitals, are exposed to excessive risk from random variation and have incentives to avoid some patients or discharge them prematurely. As evidence, critics cite individual cases in which small hospitals have incurred very high unreimbursed costs for Medicare beneficiaries which have little prospect of being balanced out by payments for treating other Medicare beneficiaries. Some critics also maintain that outliers are distributed unevenly across hospitals, creating systematic problems for those hospitals. Critics have argued that the outlier problem could be relieved by a severity measure.

Other observers have made complaints about PPS which reflect on their expectations.

4. Perceived Equity. The system is neither well understood nor perceived to be equitable by all who participate. This encourages hospitals which receive

payments lower than their costs to feel, not that they are penalized for inefficiency, but that they are cheated. Misunderstandings have also led some hospital managers and physicians to perceive outlier thresholds and even mean lengths of stay as Medicare limits on how long patients are allowed to stay in hospital.

5. Limitations for Management. The case mix adjustment system does not provide the information which hospital managers need to manage the care of Medicare patients. In particular, critics point out that the Medicare system forces hospital managers to examine costs in their hospitals, but that DRGs are not sufficiently precise for small groups of patients to allow a manager to determine whether individual physicians and departments are efficient.

6. Not Helpful for Capitation. Finally, critics most concerned with moving Medicare to capitation payment charge that case mix adjustment has been excessively focused on characteristics of individual discharges to the exclusion of the patient characteristics which must be the basis of adjusting capitation rates.

C. GOALS FOR CASE MIX ADJUSTMENT. The above criticisms, whether accurate or not, imply a reasonable set of study areas for case mix adjustment. A case mix adjustment system should:

1. Be Equitable to Hospitals. Provide to each hospital total payment which is adequate to meet the efficient treatment needs of its Medicare beneficiaries. The system should progressively clarify whether urban/rural, teaching, regional, disproportionate share, and other adjustments rest on case mix differences or on other differences among hospitals. Equity to hospitals should include appropriate protection against the cost of outlier cases.

2. Be Equitable to Beneficiaries. Pay for classes of beneficiaries and kinds of diseases an amount sufficiently related to the costs of medically necessary care so that no hospital has an incentive to discriminate against beneficiaries with certain

diseases or certain social or economic characteristics or to provide inadequate care.

3. Be Fully Understood by Hospitals, Physicians, and Beneficiaries. Those who depend on the case mix adjustment system should be able to understand it, and especially be able to understand the concepts of averaging and outlier payment on which it depends. The limitations of the system for hospital management and capitation should be explicit.

D. THE IMPORTANCE OF EQUITY. If reimbursement is not equitable, consequences could be serious both for beneficiaries and for the health care system. Medicare beneficiaries, or certain groups of beneficiaries, might find that the health care system could not afford to treat them, and beneficiaries might be denied care or dumped into hospitals of last resort. If payments are inequitable, hospitals which continued to treat all patients could face severe financial problems, not because of their own inefficiency, but because PPS did not reimburse them accurately. In the long run, the health care system might become less rather than more efficient.

Some commentators argue that there is so much flexibility and "fat" in the health care system that even fairly serious payment inequities will simply be absorbed in the short run. The high average profits reported for the first year of PPS (Chapter 4.V) support this argument, but not in all hospitals.

The designers of PPS also understood from the outset that the transition to PPS would force many high-cost hospitals to alter their practices dramatically. Conversely, no one knew how hospitals whose payments exceeded their costs would use their newfound resources. The designers of PPS used cushioning mechanisms to minimize the shock of transition by fitting the new system to historic cost patterns. In any change, there must be winners and losers, but PPS uses averages in a way which makes it difficult to determine whether a hospital loses because of inefficiency or because its patients have needs which are recognized by the clinical staffs but not by PPS. Few hospitals which lose are inclined to attribute their losses to their own inefficiency and few hospitals which gain attribute their gains to having a patient population with unusually simple needs.

Accordingly, both Congress and HCFA require a system which gives them justified confidence that hospitals win and lose under PPS primarily because of their efficiency or inefficiency rather than because of errors in the case mix system. Equally, unless hospitals, physicians, and beneficiaries believe that PPS is equitable, providers will feel justified in manipulating the system and beneficiaries may ultimately be denied needed care. This report specifically addresses these issues by exploring evidence as to which hospitals experienced substantial gains or losses in the first year of PPS (Chapter 4.V) and whether those gains or losses can be attributed to needs of patients which PPS does not currently recognize. To understand this exploration, however, the reader will need to appreciate some details of PPS, whose interdependent adjustments are quite complex.

II. THE STRUCTURE OF THE CURRENT SYSTEM

This section describes the intricate system which the architects of PPS developed to assure that per case payment to hospitals would be equitable. In Chapter 4.I, we will examine some of the policy judgments which were made in constructing this system.

A. MULTIPLE ADJUSTMENTS. PPS pays hospitals for care to Medicare beneficiaries at a per-case rate, which is set before care occurs and reflects expected rather than actual costs. The per-case rate is carefully adjusted to reflect the differences in cases treated among hospitals. The DRG system has received so much publicity that many readers assume that it is the only way in which PPS adjusts per case payments to reflect differences in case mix between hospitals. Actually, PPS uses a set of interdependent adjustments, of which DRGs are not always the most important.

PPS first determines a standard payment amount per case, which is the amount that would be paid for the average patient in the average hospital in the same kind of region (rural or urban). Currently, this standard payment is a blend of the hospital's historical cost per case, the regional average cost per case, and the national average cost per case. Under original legislation, all PPS hospitals were to be paid at a national rate starting in fiscal year 1987, but recent legislation has temporarily deferred this transition for most hospitals until 1988 (Table 1.1).

PPS makes three important kinds of adjustments to the standard payment in order to make per-case payment equitable: DRG-based measurement of case mix, outlier adjustments, and hospital-level factors. But these adjustments are made separately and in a somewhat more complex sequence: selection of a rural or urban basic payment amount, adjustment of the payment amount by the local wage index, multiplication of the wage-adjusted payment amount by the DRG weight, application of any outlier adjustment, multiplication by any indirect teaching adjustment, and, finally, addition of direct teaching and capital "pass-throughs."

Finally, some facilities are so different from the typical PPS facility that either Congress has excluded them from PPS by law or the Secretary has exempted them by regulation. Excluded facilities include hospitals with an average stay over

25 days and pediatric hospitals, hospitals and distinct units which primarily provide psychiatric or rehabilitation services, and hospitals located in States which are exempt from PPS. In addition, certain special facilities are exempted by regulation, including certain alcohol, and comprehensive cancer hospitals and units (Langenbrunner 1986). These facilities are paid under Tax Equity and Fiscal Responsibility Act of 1982 regulations which will not be discussed in this report.

B. URBAN-RURAL DIFFERENCES. PPS classifies all hospitals as urban (located in a Metropolitan Statistical Area (MSA) or equivalent area or rural (all others) and calculates standard payment amounts separately for the two groups. However, there are two further special arrangements. First, the Department has defined in regulation certain features of size and range of services under which about 160 hospitals are designated as rural referral centers and paid like urban hospitals. Second, about 360 more hospitals have been defined as sole community hospitals because no other hospitals are reasonably accessible to their patients; they will receive indefinitely a blended rate which is 75 percent hospital specific and 25 percent a regional rural rate. The Department will send Congress a separate report on urban/rural differences.

C. ADJUSTMENT FOR COST OF LABOR. The next step is to adjust the standard payment amount to reflect the cost of labor. Originally this local index was based directly on information from the Bureau of Labor Statistics but more recently the Department has developed a more sophisticated measurement which includes information from direct surveys of hospitals and controls for local variation in the number of part-time employees.

D. DRGs AND THE CASE-MIX INDEX. Once the hospital's standard payment amount is determined and adjusted to reflect wage rates, the per-case payment rate is further adjusted to reflect the patient's illness and treatment by the DRG system. All discharges are assigned to one of the DRGs (see Chapter 2.IV for details). To determine the payment for a specific case, PPS gives a DRG-specific weight (published annually in the Federal Register) to each case, and the weight is multiplied by the standard payment amount for that hospital. The average DRG weight of a hospital's Medicare discharges is the hospital's case mix index (CMI). Thus, while payments are actually made on a case-by-case basis, the average payment across all cases is equal to the hospital's standard payment amount multiplied by the CMI and further adjusted as described in Parts E and F of this section.

DRGs address severity in several ways--by the use of specific disease codes, by differentiation among procedures, by the use of age, and by taking certain secondary diagnoses as indicators of "comorbidity and complications." But DRGs do not use any clinical data or laboratory findings to measure need for care and the adequacy of the way DRGs measure severity of illness remains unproven. A number of systems have been proposed as possible refinements to or replacements for DRGs (Chapter 3).

E. OUTLIER CASES. After adjusting for DRG, PPS makes special additional payments for "outliers"—patients with unusually costly or unusually lengthy stays. The Social Security Amendments of 1983 required that 5-6 percent of payments under PPS be made through a system designed to account for outlier cases. Outlier payments work as follows:

a. Day Outliers. For each DRG, PPS sets a threshold length of stay for outliers; in 1984 the threshold was defined as the lesser of 20 days or 1.94 geometric standard deviations above the geometric mean length of stay. All necessary inpatient days above the stay threshold are reimbursed at a flat per diem rate which is 60 percent of the PPS payment per day for patients assigned to that DRG (defined as the PPS payment divided by the geometric mean length of stay).

b. Cost Outliers. In addition, the outlier policy makes allowance for patients who have very high bills but do not exceed the length of stay threshold. For these "cost outliers," PPS pays about 43 percent of charges exceeding a threshold (which was $12,000 or 150 percent of the DRG rate in 1983-1984); 43 percent results from multiplying the 60 percent marginal cost figure by 0.72, which is the average ratio of operating costs to charges in Medicare bills. The thresholds were calculated so as to have 15 percent of cases cost outliers and 85 percent day outliers. Chapter 4.VI examines the outlier problem in detail.

F. INDIRECT TEACHING COSTS. The teaching adjustment is applied to the total calculated payment, including outlier payments. The original increase was 11.6 percent for each tenth of an intern or resident per hospital bed, but the COBRA reduced this adjustment and incorporated it into a more complex payment formula. Table 1-2 compares payments under the old and new formulas. This payment was proposed as support for indirect costs of medical education and is not explicitly designated as related to the case mix of teaching hospitals.

G. REGIONAL DIFFERENTIALS. Regional adjustments were designed to cushion hospitals during the transition period (1983-1987) to fully national rates; they are scheduled to be phased out.

H. DIRECT TEACHING AND CAPITAL PASS-THROUGH. PPS places the direct cost of approved medical education activities and the costs of capital outside of the prospective, case-based payment system. These costs, subject to certain limitations, are "pass-throughs," and Medicare's share of these costs is paid directly to the hospitals. This report will not examine these issues.

I. THE INTERACTION OF ADJUSTMENT FACTORS. The preceding discussion makes clear that the Medicare case mix measurement system is interdependent with PPS payment rules in ways which make evaluation difficult. PPS determines payments not only by DRGs but also by the outlier payment system and hospital-level adjustments. For example, if urban/rural or teaching adjustments are inaccurate, separating the results from inaccuracies in the case mix and outlier systems will be very difficult. Likewise, evaluations of the DRG classification system will substantially underestimate the applied power of the DRG system within PPS if cases paid as outliers are included in the evaluation but the outlier payments are omitted. A comparable evaluation error will occur if the outlier system pays less than intended and the underpayment is mistaken for a defect in DRGs. Finally, assessing the outlier system depends on a consistent interpretation of whether it is intended to complement and extend the DRG system by paying full costs or is intended to create strong incentives for shortening stays by paying less than costs.

There are no studies in the literature that take full account of all these adjustments in evaluating the adequacy of DRGs or the equity of PPS (Cretin and Worthman 1986). Even Cretin et al.'s (1986) study, which was the first comparison of actual payments with costs under PPS, had to create "synthetic" costs because 1984 Medicare cost reports were not yet available for analysis. We believe that the current report is the first national comparison of actual PPS payments with actual PPS costs.

III. EQUITY AND PAYMENT DISCREPANCIES

A payment discrepancy is the difference between the payment which Medicare makes and the actual cost of care. Individual payment discrepancies do not necessarily produce unfairness--for example, an overpayment on one case may be offset by an underpayment on the next case, producing an equitable average payment to the hospital. Thus, the implications of individual payment discrepancies in a payment system built on averages can be subtle and very difficult for a hospital manager to assess.

PPS legislated a profound change in hospital payment which is not yet universally understood by either PPS critics or hospital managers. Medicare no longer pays the cost of a specific case, or the cost of comparable cases treated in an average hospital, or even the cost of treatment determined to be necessary for an individual case. Instead, PPS seeks to pay hospitals fairly for the total care of their Medicare patients.

Because PPS uses hospital-level factors based on averages, a finite number of DRGs, and an outlier system based on average marginal costs, it cannot pay an exactly correct amount for each case. The underlying assumption in PPS is that the treatment needs of individual patients in a DRG do vary from the average of needs for all patients in that DRG, but that this variation is randomly distributed from hospital to hospital. Although some cases will require fewer resources and others will require more, the average for each hospital is expected to approximate closely the average across all hospitals.

Because we cannot determine whether case mix measurement is accurate by looking at individual cases, we must distinguish random cost variations from systematic cost differences between hospitals (McGuire 1985). The next three sections examine random discrepancies, systematic discrepancies, and patient-oriented discrepancies.

A. RANDOM DISCREPANCIES. Random discrepancies result from random variation among cases and are the necessary result of paying average costs. With a sufficiently large number of admissions, random discrepancies average out and are unimportant for the hospital's budget (this protection is known as the law of large numbers); all but the smallest hospitals have enough admissions to secure this protection, but many individual departments, physicians, and DRGs do not. If a DRG has 25 cases and a typical amount of variation, the average cost will differ from the national average cost (and therefore from payment) by at least 10 percent more than half the time and by at least 20 percent more than a fifth of the time.

About 1 percent of hospitals with 35 beds and a 75 percent occupancy rate would experience a 10 percent shortfall by chance alone.* But, since such shortfalls are due to chance, they cannot be expected to persist from year to year, and only one hospital in 10,000 could expect such an experience two years running. Random payment discrepancies attributable to within-DRG variation are very small for classes of hospitals because the law of large numbers affords protection proportional to the total number of admissions in the entire class of hospitals.

Reducing the amount of variation within a DRG or increasing the precision of outlier payments protects hospitals against random risk even if there is no systematic risk. Even if overall payment is equitable, random payment discrepancies can create significant problems for hospital managers, not only because they can create cash flow problems but also because they make it difficult for the manager to determine whether the hospital is operating inefficiently or whether it is experiencing a chance run of difficult patients. This is especially so because management can rarely be conducted on a hospital-wide basis. Rather, individual departments and even units within departments must be managed, and these have fewer discharges and are, therefore, more subject to chance fluctuations. Thus, although Medicare payment may be equitable if the average payment to hospitals is equitable, hospital managers now need to assess the appropriateness of costs for smaller units in order to manage effectively. This dilemma leads managers to complain that DRGs are too imprecise (Chapter 2.I.C explores this issue) and may lead them to adopt stringent cost-saving strategies in response to a chance run of expensive cases. Other managers, who appear not to understand the concept of averages, describe every patient with costs above average as a "loser" without examining the counterbalancing profits on patients with costs below average.

---

*If the average coefficient of variation for each DRG is 0.82 (HCFA unpublished data), the relative standard error is about 4 percent for payments to hospitals with 400 Medicare discharges a year. A hospital with 35 beds, 30 precent Medicare discharges, and a 75 percent occupancy rate would expect this protection.)

B. SYSTEMATIC DISCREPANCIES. Systematic discrepancies are those against which the law of large numbers affords no protection. The remainder of this report focuses on possible causes of systematic discrepancies, including unmeasured variation in treatment needs among patients in different hospitals, inaccurate hospital-level adjustments, inaccuracies in outlier payment policies, and differences in efficiency among hospitals.

To the extent that PPS is increasingly able to explain and correct systematic discrepancies between hospitals, the problems posed by random discrepancies will diminish because hospital managers will be able to distinguish random and systematic discrepancies and to have increasing confidence that random variations will average out over time. A reasonable measure of success will be the ability of PPS to produce DRGs or other classification methods which have increasingly homogeneous costs across hospitals regardless of variation within hospitals.

C. POTENTIAL DISCRIMINATION. The simple distinction between random and systematic discrepancies ignores the incentive for a hospital to identify discrepancies which have been randomly distributed in the past, to send unprofitable cases elsewhere (dumping), and to recruit profitable cases (skimming).

Hospitals can do this if they can predict that the PPS payment will differ from the cost of necessary treatment for any particular patient and can find a way to fend off losers and recruit winners. Thus, variation or heterogeneity within a DRG creates risk of a discrimination against unprofitable patients if and only if a hospital can, or believes it can, identify those patients who will be "losers" either in advance of admission or soon after admission.

A troubling implication of this analysis is that payment to the hospital need not be incorrect in order to create discrimination. If, for example, hospital administrators commonly believe that admissions from nursing homes or admissions of impoverished beneficiaries are "losers," those beneficiaries may encounter discrimination even if the common belief is in error. Thus, determining whether such incentives exist is important so that the incentives can be corrected if they are found and so that hospital managers can be educated if there is no rational ground for discrimination (see Chapter 2.111.A and 3.VII.).

A second implication is that, if the more costly patient can be identified before admission, even a well-designed and equitable payment system can become the basis of discrimination for a manager more interested in profits than in providing health care. Peer Review Organizations (PROs) have a special role in identifying such problems if they begin to emerge and a special mandate to prevent the "dumping" of patients through premature or inappropriate discharge.

D. HOSPITAL DIFFERENCES AND EQUITY. A critical question in determining equity to beneficiaries and hospitals is whether, after case mix correction and hospital-level adjustments, hospitals either do or potentially can receive, select, or recruit patients whose treatment costs differ systematically from the average patient in the same DRG.

This question can be put in a different way. If we examine all patients in a region who are assigned to a DRG, is the variation in treatment needs within that DRG randomly distributed across hospitals or is it systematically associated with factors which either do or could determine the hospital which a patient enters? If variation is correlated with hospital selection, then some hospitals may get patients with greater treatment needs, others may get patients with less treatment needs, and the system is potentially unfair or biased. If the hospital-level adjustments and outlier policy do not correct the bias, then the system is unfair in practice.

Teaching hospitals, public hospitals, and specialty hospitals can easily demonstrate that their patients have distinctive characteristics (e.g., being referred, being poor, having special kinds of illness). The issue is whether these distinctive patient characteristics are correlated with the variation in treatment requirements within a DRG. If that correlation exists and if hospital-level and outlier adjustments do not correct for it, then payments to these hospitals will not reflect their patients' needs.

PPS was constructed with a built-in conflict between an effort to assure through adjustments that groups of hospitals did not win or lose and an implicit assumption that some hospitals were less efficient than they should be. The conflict can only be avoided by assuming that all classes of hospital are equally efficient, with variations in efficiency occurring exclusively within classes but never between classes. Despite this paradox, the occurence of major systematic winners or losers under the first year of PPS would raise questions about the equity of the system.

E. SUMMARY. Section I described the goals of case mix adjustment as equity to both hospitals and beneficiaries and defined equity as paying the cost of necessary care. Section II outlined the way in which PPS seeks to accomplish those goals. Section III underlined the difference between paying fairly for each case and paying fairly on average and emphasized that the policy of paying fairly on average poses difficulties for both hospital managers and PPS managers who seek to determine whether the system is indeed equitable. The next chapter examines the problem of predicting necessary care for individual patients and examines the success of the DRG system in meeting this goal.

Table 1.1

Transition from Hospital-Specific to National Rates

A. Hospital-specific and Federal Portions as percent of total payment

	Hospital Fiscal Year Begins				
	10/1/83-9/30/84	10/1/84-9/30/85	10/1/85 9/30/86	10/1/86-9/30/87[1]	after 10/1/87
Hospital-specific	75	50	50-45[1]	25	0
Federal	25	50	50-55[1]	75	100

B. Regional and National Components of Federal Portion as percent of Federal portion

	Discharges Occurring				
	10/1/83-9/30/84	10/1/84-9/30/85	10/1/85 9/30/86	10/1/86-9/30/87[1]	after 10/1/87
Regional	100	75	75	50	0
National	0	25	25	50	100

[1] The blend remains 50 percent hospital-specific through the first 7 months of this reporting period and then becomes 45 percent hospital-specific for the last 5 months of the reporting period.

Note: Hospitals in Oregon move to national rates somewhat more rapidly than this table shows.

Source: Federal Register 1986; 51:16774.

Table 1.2

Indirect Medical Education Adjustments under Old and New Law
Percent Increase in Payment

Residents per 100 beds	Discharges Occurring		
	Before 5/1/86	5/1/86-9/30/88	after 10/1/88
none	0	0	0
10	11.6	8.1	8.7
25	29.0	18.9	20.7
50	57.9	35.7	39.7

Source: Social Security Amendments of 1983 (P.L. 98-21) and Consolidated Omnibus Budget Reconciliation Act of 1985 (P.L. 99-272).

CHAPTER 2

PREDICTING NEEDED CARE FOR INDIVIDUAL PATIENTS

Understanding the issues of severity of illness, intensity of care, and outlier payment policy requires understanding the basic problem of predicting need for care from facts about the patient and the patient's illness. This chapter focuses on using a set of patient characteristics (the independent variables) to predict a patient's treatment needs (the dependent variable). Section I examines the assumptions which Medicare built into PPS by adopting DRGs and by the way PPS pays for outliers, Section II explores the concept of necessary care and looks at the sources of variation for which Medicare should and should not pay, Section III reviews the issues in choosing variables to predict necessary care, and Section IV assesses DRGs as a case mix measurement system. Chapter 3 evaluates alternatives to DRGs using the framework applied to DRGs in Section IV, and Chapter 4 examines the problem of predicting the average care needs for all patients in a hospital.

I. THE PPS MODEL OF DIFFERENCES IN TREATMENT NEEDS

In Chapter 1 we reviewed the structure of PPS. We now examine both the DRG system and the outlier payment system in order to understand the assumptions about necessary care for individuals which are built into PPS.

A. THE MODEL IMPLICIT IN DRGs. The DRG system operates from the Uniform Hospital Discharge Data Set (UHDDS), which hospitals collect for every Medicare discharge and for most other discharges. UHDDS contains only the principal diagnosis, four additional diagnoses, three procedures, the patient's age and discharge status (alive, dead, against medical advice, transferred to another hospital, transferred to another institution, sent home), admission and discharge dates, and expected source of payment. Thus, Medicare's decision to adopt DRGs meant that PPS could use only UHDDS variables as indicators of need. Adopting DRGs also meant incorporating into PPS certain conventions which had been accepted by the team which created the DRGs.

-- Death would be used to classify patients only if it resulted in lower resource consumption; Medicare would not reward hospitals if patients died.

-- There would be only 300-500 DRGs; the number of groups would be small enough to assure significant numbers of patients in each group in most hospitals. This prevented creating small DRGs for unusual cases and meant that a special policy would be necessary to pay for such cases if Medicare decided that they were not evenly distributed or that they presented an unacceptable risk to hospitals. This decision reflected the origin of DRGs as a hospital management studies tool which needed a meaningful number of cases in each category within a hospital (Cretin and Worthman 1986).

-- A single set of comorbidity and complication indicators would be used for all DRGs; testing separate lists for many different conditions would have required computer resources and personnel which were not available. This meant that the system would not seek to distinguish between complications and comorbidity and might be criticized for rewarding care resulting in complications.

-- Surgery would be used to make the first division of groups within a major diagnostic category; since the groups were intended to be clinically homogeneous, patients were separated according to the most important division in medicine. This meant that the system would accept the surgeon's decision to operate as a reasonable basis for deciding payment; clinical decisions with more modest cost implications were generally not recognized as a basis for payment.

The decision to limit the number of groups and the decision that DRGs would not be designed to pay outliers led directly to creation of an outlier policy to safeguard payment equity for hospitals and individuals.

B. OUTLIER POLICY. The most important ways in which outlier policy incorporated an implicit model of interpatient differences was the way in which marginal costs were defined and the decision that PPS would not have an outlier policy for low-cost cases.

-- Marginal costs were calculated on an assumption that the marginal cost for outlier cases was significantly lower than the average per diem cost for nonoutliers. Information to test this hypothesis was not available (Chapter 4 provides recent evidence on the issue).

-- The legislative decision not to have short-stay outlier policy meant that unusually short-stay patients were not treated as qualitatively different from usual patients. Although this supported a program emphasis on incentives for efficiency, it missed an opportunity to pay differently for patients admitted for observation of possible disease rather than for treatment.

II. DEFINING NECESSARY CARE

The acid test of an equitable payment system is that it pays the cost of efficiently, providing necessary care. But defining what care is necessary is not simple and becomes especially formidable when the payment is based on the average cost of comparable cases rather than on retrospective review of whether the care actually rendered was appropriate.

A. THE SEVERITY OF ILLNESS MODEL. To explain the difficulty of projecting necessary care from information about the patient, we will start by considering the "severity of illness" model. This model rests on the assumption that we can classify patients into categories which are homogeneous with regard to all characteristics, except severity of illness, and that we can then determine the cost of necessary care if we know the patient's "severity of illness." When treatment costs vary among patients with the same severity of illness in the same category, the severity model attributes those differences primarily to differences in efficiency of care and does not make allowance for other sources of variation. Although this model is more simplified than that used by any investigator in the field, it highlights important issues.

1. Defining "Severity of Illness". To understand this model we need first to define severity of illness, whose meaning varies in four ways.

a. Professional Perspectives. To the physician caring for an acutely ill patient, clinical severity of illness is usually a measure of the likelihood of irreversible loss of function, failure of an organ system, or death. To nurses caring for the patient, and to family and social service agencies, severity may also denote the patient's total chronic impairment and dependency on others. To a hospital administrator, severity may mean how many resources will be needed for a patient's care.

b. Response to Treatment. Some investigators include the patient's response to treatment in measuring severity. This definition implies that the "true" severity of an illness is reflected in whether it is easily treatable. If two patients with pneumonia have identical clinical findings on admission to a hospital, the one who fails to respond promptly to treatment is, from this perspective, more severely ill than the one who responds promptly. Unfortunately, if we measure severity after the beginning of treatment, we risk confounding patient condition with the effectiveness of care rendered. Most investigators, therefore, agree that a system which could predict response would be preferable to one which required actual observation of the patient's clinical course, but such precise predictions are beyond the state of clinical art and far beyond the state of computer based systems based on abstracted data. The next chapter describes the APACHE system, which was designed to forecast mortality.

c. Chronic v. Acute Severity. Some investigators focus on disease progression as a chronic condition (e.g., Disease Staging, which is based on the model of chronic diseases such as cancer). Others focus on the patient's acute condition at the moment of hospital admission (e.g., APACHE and MEDISGRPS). The two models clearly imply different strategies for description and measurement.

d. Treatable v. Untreatable. Theoretically we might wish to define severity as the portion of the patient's illness which is amenable to treatment thus excluding from payment those components of severity which are not treatable. Such an approach implies kinds of triage which are beyond the current state of the art and may not be desirable in some cases, but the idea is relevant to the severity model.

e. A Definition. In this report, severity of illness is the patient's clinical condition at the time of hospital admission as indicated by the clinical likelihood of adverse outcomes such as death, permanent impairment, or disability.

2. Limitations of the Severity Model. Even if we could agree on a definition of severity, the severity model is limited by variation on four important dimensions:

--accepted medical practice,

--the reasons why patients who have the same condition are admitted,

--variation in what patients want and need from their health care, and

--the effects of death on costs.

These limitations have significant payment implications and lead us to the part of the congressional mandate which concerns intensity of care.

B. THE INTENSITY OF CARE OR VARIATION IN PRACTICE PATTERN MODEL. A major limitation of the severity model is the implicit assumption that there is one way to treat a patient with a particular condition at a particular level of severity. This implicit assumption suggests that there is consensus on a common set of procedures or a common set of outcomes which can be expected. In fact, procedures, expectations, and outcomes vary widely, resulting in substantial variation in intensity of care for patients with similar severity of illness.

1. Practice Pattern Variation. The most dramatic evidence of variation in medical practice describes the actual frequency of hospitalization. Wennberg and Gittlesohn (1982) and Wennberg et al., (1984) have shown that rates of hospital admission for various procedures vary widely over small areas; Chassin et al. (1986) have shown comparable variation over large areas. These variations have not been shown to reflect either inappropriate medical practice or inefficient use of resources. Hospitalization rates are sharply lower (Luft 1978, 1981) for patients in prepaid health plans compared to those in fee-for-service settings, and this is not explained by selective enrollment (Manning et al. 1984). Since PPS pays only for patients who have been admitted, differences in admission practices might not be critical to cost per admission, but if the patients needing less care are not admitted, average costs of other admissions may be higher. Gornick (1982) has provided some evidence that this reciprocal relationship occurs.

Mitchell et al. (1985b) provide a useful example of practice pattern variation by contrasting treatment for similar conditions in four states. Cases which fall within the same DRGs vary significantly across States (an issue of considerable importance when we consider the relation of case mix evidence to national rates in Chapter 4), but the differences in practice patterns appear to be far greater than variation in case severity. Iezzoni and Moskowitz (1984) used different methods to reach parallel conclusions in comparing New Jersey and North Carolina data; Cromwell et al. (1985) show similar variation across hospital types. Variation in practice patterns across regions is well-documented, and the shorter stays observed in western States compared to eastern States is well known (Chassin 1983, Beebe et al. 1985). At least one careful study (List et al. 1983), has shown that this variation results from practice pattern variation rather than patient characteristics.

The problem of practice pattern variation was familiar to the architects of PPS, and the decision to phase in national rates reflected a considered conclusion that, while uniform practice patterns were not necessary, it was unreasonable to continue to pay different amounts in for essentially comparable cases in different regions. But such variations in practice patterns conflict with the severity of illness model because heterogeneity of practice does not support the assumption that there is a single medically-accepted way of treating a patient with a given condition and a given severity level. Practice pattern heterogeneity also makes determining whether patients are truly comparable very difficult.

Whether Medicare should pay for different practice patterns is a difficult question. Few would argue that Medicare should pay more than the lowest cost if identical patients are treated at different costs with identical outcomes; Medicare should not pay for therapeutic activity which has no benefit to the patient. But there are few areas of medicine where patient care can be based on unequivocal outcome studies. The picture is further obscured because variations in reason for admission and goals of treatment are often mistaken for variation in practice styles.

2. Variation in Reason for Admission. Patients with identical conditions and identical severity of illness may be admitted for sharply different reasons. The admitting physician's intent is often independent of clinical condition and reflects such issues as:

a. Diagnostic v. Therapeutic Intent. Some patients are admitted to a hospital after the diagnosis is established while others are admitted for diagnostic evaluation. The procedures performed can indicate this difference in some cases, which may affect DRG assignment. For example, if a cystoscopy is not performed during an admission when a patient has a prostatectomy, it is likely that one was performed prior to admission and that the diagnosis was established before admission. In the case of coronary artery bypass grafting, patients are assigned to different DRGs according to whether a cardiac catheterization was performed during the same admission. A catheterization suggests that the diagnosis was not fully established before the admission, but absence of a catheterization does not indicate whether the diagnosis was established prior to admission or by other procedures during the admission. In general, emergency admissions are likely to require inpatient diagnostic evaluation and to have higher costs (Munoz et al. 1986).

b. Complete v. Staged Treatment. The example of staged treatment which has been most intensively discussed for DRGs is the contrast between two admissions and one admission for bilateral major joint surgery; this problem has been resolved by creating a new DRG for multiple procedures. But similar problems exist for many other conditions, such as whether to operate on a patient whose cholecystitis has been treated or whether to return that patient for surgery at a later date (Gertman and Lowenstein 1984).

c. Palliative v. Curative treatment. The choice between palliative and curative treatment is one of the most sensitive in clinical medicine, involving subjective and value-laden judgments. Some patients wish heroic measure;

others do not. Bedell and Delbanco (1984) found, for example, that 8 of 24 patients who had been successfully resuscitated after a cardiopulmonary arrest said that they had not wished resuscitation and would not wish it in the future. Patients choosing between alternative cancer therapies make similar decisions. McNeil et al. (1982) have shown that patient decisions are highly sensitive to the nuances of the way the physician presents the decision. Issues of this kind appear to be beyond the reach of any severity measurement system.

d. Operative Procedures. Objective data tend to capture well whether a patient should be treated surgically or medically, which is one reason why surgeons rather than computers make such decisions. Even if such data capture were possible at a point in time, the criteria would probably be dated by the time they could be put in place for Medicare payment. Here we note only that predicting whether surgery is necessary from admission severity appears impractical. In Section III.B, we will discuss the problems raised by determining payments from procedures performed.

3. The Effects of Death on Costs.

a. Patients Who Die In Hospitals. A number of studies have now shown that the most seriously ill patients do not necessarily consume the largest amount of resources--some die relatively quickly and receive little treatment (Conklin et al., 1984a, Conklin 1985; Horn et al., 1984b; Wagner and Draper 1984). The DRG system recognizes this effect in myocardial infarction, where a special DRG for patients who die receives somewhat lower payment than DRGs for patients who live. If death could be predicted accurately, this effect would not confound the effort to predict costs from severity of illness, but these severely ill patients tend to be extremely expensive if they live, causing serious predictive problems.

b. Where Patients Die. Many patients do not die in hospitals because they, their physicians, or the health care system elect to discharge them to other settings before they die. Hospices are only one example of health care system arrangements which promote such discharges. Patterns vary widely from State to State (Chapter 4.III) and presumably reflect both patient preferences, and physician judgment as to the circumstances which can best combine good medical care with a supportive environment.

4. Severity and Cost of Care. These reservations regarding the severity model may appear somewhat academic to readers who are convinced that, nevertheless, severity differences obviously account for most variation in cost of care. Although available data do suggest that severity of illness is an important source of variation in cost per case within hospitals (Horn and Sharkey 1983; Horn et al., 1984b; Horn et al., 1985a,b; Brewster et al., 1985b, Wagner and Draper 1984) and between hospitals of a similar type (Horn et al., 1984, 1985a,b), there is little evidence that severity differences explain the majority of differences in costs between hospital types. In Chapter 3. VII, we will see that such "obvious" predictors of severity as poverty, disability, extreme age, and previous hospitalization do not, in fact, predict higher costs once DRG and hospital are taken into account.

C. SOURCES OF VARIATION IN COST OF CARE FOR WHICH MEDICARE DOES AND DOES NOT PAY. In principle, Medicare pays only for causes of cost variation beyond the hospital's control. To clarify the concept of needed care, we will consider seven sources of cost variation. Of these, Medicare agrees to pay for three, seeks not to pay for one, and has no explicit policy regarding three:

1. Accepted Sources of Variation. Medicare accepts responsibility for:

a. Factor Prices. These are the prices which the hospital must pay for labor, capital, and house staff services. As noted above, however, Medicare implicitly assumes that prices of supplies and equipment, and total costs for energy and security are uniform across hospitals.

b. Referral Patterns. Patients are referred to hospitals and transferred between hospitals so that some hospitals have case mixes which are different from those of other hospitals. Some of this variation is well-described by DRGs (e.g., specialization by some hospitals in certain kinds of surgery), but DRGs may not capture costs when referrals result from diagnostic difficulty or severity of illness, which may be the case for teaching and referral hospitals.

c. Demographic Variation in Patterns of Disease. Medicare would accept variations in costs which result from differences in kinds and severity of illness across hospitals and regions.

2. Rejected Sources of Variation. Medicare seeks not to pay for:

a. Inefficiency of Care. One of the major purposes of PPS is to give hospitals incentives to become more efficient. Medicare hopes that PPS will specifically not pay for differences in efficiency. Inefficiency includes unnecessary admissions, unnecessary inpatient days, and unnecessary services during an admission as well as other kinds of inefficiency.

3. Ambiguous Issues. Medicare policy is ambiguous on paying for:

a. Variation in Available Support Systems. Medicare seeks not to pay for hospital care which is required because of the absence of community resources such as nursing homes; to pay for such care would encourage communities inefficiently to substitute inpatient services covered by Medicare for less expensive services which are often not covered. On the other hand, Medicare does pay for "administratively necessary" days of care and this care was built into DRG weights and payments.

b. Admission Threshold Variation. Medicare clearly refuses to pay for unnecessary admissions. But we have noted that the threshold for necessary admissions appears to vary among physicians and across regions and that this variation may affect both total costs and cost per admission. PROs review admission thresholds, but there is no national standard.

c. Practice Variations. There are differences, well within the range of acceptable medical practice, in what different physicians do for patients with identical conditions. We do not know the extent to which variation in treatment activity reflects differences in goals, such as cure versus palliation or diagnosis versus treatment. As noted, Medicare would seek not to pay for practice variations which have no effect on outcome.

D. USE OF ACTUAL PAYMENTS. In practice there is no gold standard for measuring need for care, and investigators have widely accepted the use of actual cost of care as a kind of silver standard. Only a very small number of studies have sought to go beyond this standard (Young 1985; Strumwasser and Pararipe, 1985; Systemetrics 1983; Kay et al., 1986; Gertman and Bucher 1971; Gertman and Restuccia 1981; Restucccia et al, 1984) and none has examined whether classification systems predict necessary care as well or better than they predict care rendered. Indeed, many investigators have been unable to obtain cost information and have had to settle for charges or actual length of stay. Although Cotterill et al. (1985) have shown that charges can be accurately substituted for costs in some cases, charges cannot be used to assess whether payments reflect costs. For the remainder of this report, we will, therefore, be dealing almost exclusively with costs of care actually rendered rather than with necessary care, but the reader should keep in mind that this is a compromise. The importance of this distinction will become clearer in Chapter 4.I when we discuss the kinds of cost variations for which PPS does not pay because they are thought to be unnecessary.

III. PREDICTORS OF THE COST OF NECESSARY CARE: THE INDEPENDENT VARIABLES

In this section we examine four issues in choosing variables to predict the cost of an admission or to determine the amount of a payment:

- The implications of variable choice for medical practice,
- The choice of predictive variables versus variables which describe what was done,
- The administrative feasibility of using different variables, and
- The basic problems in design of a system which fully specifies necessary care.

At the end of Chapter 3, after examining data on alternative case mix measures, we will consider what variables are most credible.

A. THE CONCEPT OF NEUTRALITY: HOW THE CHOICE OF PREDICTOR VARIABLES CAN SHAPE CLINICAL PRACTICE. The choice of predictor variables in a case mix system is extremely important in determining not only the power and administrability of a system but also its impact on clinical practice. The choice of independent variables expresses the philosophy of the system as to what sources of variation in cost of care are legitimate, and the choice directs a system clinically and socially. A key concept in the analysis which follows is that including a variable in a case mix measurement system tends to make the system neutral with regard to that variable.

1. Fiscal Neutrality to Clinical Decision-Making. A system which neither rewards nor penalizes a hospital for performing a procedure is fiscally neutral to that procedure. For example, DRGs use major procedures to classify patients. If reimbursements are properly calculated for the DRGs with and without surgery, the difference in reimbursement is the cost of surgery, all other things being equal.

Thus, the hospital is neither rewarded nor penalized for performing surgery and the provider is free to make this decision on clinical grounds. Coffey and Goldfarb (1984) report that, while teaching and nonteaching hospitals in Maryland have very similar case mix as described by Disease Staging, teaching hospitals have more complex case mixes as described by DRGs. This difference occurs because teaching hospitals do more procedures for patients at a given stage of illness, and DRGs use procedures to classify a patient while Staging does not. One might argue that this increased frequency of procedures represents overutilization and should be discouraged or one might argue that it represents appropriate concentration of specialized procedures in specialty centers and should be treated with neutrality. If one adopted the former view, one would (if Staging and DRGs are otherwise equal) adopt Staging for case mix classification because Staging, by not using procedures to classify patients, creates an incentive for hospitals not to incur the costs of surgery. If we adopted the latter view, we would use DRGs for case mix classification since DRGs are neutral to performance of procedures and, therefore, do not penalize a hospital for becoming a surgical referral center.

The same argument can be applied, for example, to whether hospitals treat psychiatric patients in more expensive, specialized units. If such treatment is useful, PPS should consider such care a procedure and use it to classify patients for payment. If special units are wasteful, PPS should not use care in special units for classification. This would create strong economic incentives to hospitals not to operate specialized units. These two examples suggest that the problem of neutrality is a far-reaching issue since it raises the problem of the extent to which Medicare will pay for treatments which are more expensive but more effective, for more expensive treatments whose efficacy is unproven, and for patient and physician preferences. To the extent that higher intensity of care is not recognized by the payment system, intensity of care is likely to diminish.

2. Neutrality to Choice of Patients. Chapter 1 identified equity to beneficiaries as a critical element in the equity of PPS. To assure such equity, PPS should give providers no incentive to prefer, or to avoid, certain classes of beneficiaries. Such incentives might exist if, for example, admissions from nursing homes were more expensive than admissions from other sources. The incentive could be neutralized by making source of admission a variable in determining payment, so that when such increased costs were predictably present they would be recognized by appropriate payments (see Chapter 3.VIII.D).

3. Neutrality to Diseases. Neutrality to diseases across classes of providers may be a significant issue. Cretin et al. (1986) provide evidence that DRGs which are profitable in some classes of hospitals may be unprofitable in others. If differences in case mix or in appropriate treatments explain these differences, the resulting incentives may be undesirable .

4. Neutrality Regarding Exceptional Cases. A system which uses outlier status as a payment variable also becomes more neutral toward these exceptional patients. To the extent that Medicare wishes hospitals to be willing to take on exceptionally expensive patients, an outlier payment policy is useful. But to assure true neutrality, such a policy might need to pay a larger portion of costs than would be consistent with the incentives of PPS. Chapter 4.VI examines this problem in detail.

B. ASSESSING PATIENTS AT ADMISSION v. ASSESSING THROUGHOUT THE STAY.

Differences in time perspective in the definition of severity of illness contribute to an important controversy in predicting necessary services: can a prospective payment system legitimately use clinical events, procedures, or actual costs and charges occurring during the course of a hospitalization as indicators of patient needs?

1. Using Clinical Events. The use of events occurring during hospitalization to make case mix adjustments, and ultimately to determine payment, arouses intense feelings among program planners and theorists for two reasons:

- Using events occurring during hospitalization to determine payment suggests a reversion to cost based reimbursement, and
- Considering patient condition after admission as a basis for payment suggests rewarding ineffective care.

From this perspective, it is useful to realize that PPS already considers such data in three important ways:

- Operative procedures are used for making the most important classification decisions in the DRG system, and there is evidence that they account for most of the explanatory power of the DRG system;
- Complications occurring during hospitalization are used to make further classification decisions; in general, PPS pays more for patients who develop complications; and
- Law prescribes paying marginal cost beyond the outlier threshold.

To understand the implications of using intrahospital events to determine payments, we should distinguish between paying for the performance of clinical activities, such as surgery, and using certain procedures as indicators that the patient is very ill and merits higher payment.

a. Paying for Surgery. The fact that surgery is performed can permit a significant number of inferences about the patient if the principal procedure is typically elective:

- the diagnosis was known prior to admission
- the patient's condition on admission was good and suitable for elective surgery

-- the purpose of the admission was to perform surgery

-- the resources needed will be largely dictated by the procedure performed.

Consistent with this, a number of investigators have found that surgical DRGs have far less internal variation than nonsurgical DRGs (Mitchell et al. 1984, 1985a,b; West et al. 1985; Frank and Lave 1985). The procedures performed can also provide evidence of the variation in therapeutic intent among patients who may appear identical at admission, a kind of variation which severity of illness measurements are unlikely to capture. This strategy also provides neutrality to performance of surgery, which is important for surgical referral centers. Since reviewing the clinical appropriateness of major surgery is feasible, paying for surgery provides a substantial reward in accuracy at the cost of a rather modest step toward cost-based reimbursement. Using operative procedures as a measure of need is, however, troublesome because wide regional variation in frequency of procedures raises a question about the relation of actual procedures to necessary procedures (Chassin et al. 1986; Wennberg et al. 1984; McPherson et al. 1982).

b. Paying for Indicator Procedures. A more serious hazard rises when procedures are taken as indicators of severity and the payment rises far more than the cost of performing the procedure. This can create a significant incentive to the hospital to perform unnecessary procedures because the payment is "leveraged," and certainly provides no incentive to restrain use of these "indicator services." Another commonly cited risk is that a system which considers care actually rendered rewards poor care which results in resource-consuming complications. Although this is at least theoretically true, there are so many pressures to avoid such complications (including malpractice litigation, hospital quality assurance systems, and PROs) that it may not be strategically fruitful to attempt to address this problem through the payment system. On the other hand, rewards for inefficient care are a legitimate source of further concern, and a system which takes failure to improve as evidence of severity may have misclassified ineffective treatment, which should be penalized rather than rewarded.

3. Using Actual Costs or Charges. An important approach to case mix classification is the use of actual costs or charges. PPS uses actual charges for medically necessary services, discounted to approximate marginal costs, to determine payments for cost outliers (Chapter 4). The outlier system illustrates how actual charges or costs can be used without destroying the basic incentives of PPS: the substantial loss a hospital incurs before reaching the outlier threshold preserves the incentives. Nevertheless, even after PRO review of medical necessity, actual costs or charges remain only estimates of efficiently provided necessary care.

4. Use of Services and Equity. Determining payment from services rendered involves a tradeoff between two kinds of equity. On the one hand, considering actual services increases equity for unusual cases which cannot be predicted and for hospitals receiving difficult referrals. On the other hand, using actual services risks overpaying hospitals which practice inefficient or counterproductive medicine.

C. THE CONCEPT OF A "FULLY-SPECIFIED" SYSTEM. A "fully-specified" system would use all available medical knowledge to define necessary services for a condition and would then determine payment from the cost of efficiently providing these services. Attractive though such an idea may be, even the goal presents three important problems:

1. System Development. Developing and testing such a system might well demand a larger health services research establishment than can be expected to exist in the United States during the duration of PPS.

2. System Maintenance. Keeping the DRG system current with technology already puts unprecedented demands on HCFA's rulemaking process; extending this process to the full range of medical care would multiply these problems. In addition, continual modifications of data reporting rules would put severe administrative strains on everyone involved with HCFA patients.

3. Appearance of Regulation. DRGs are not, and were not intended to be, prescriptions for length of stay or value of services to be provided for each patient. But some hospital managers have nevertheless interpreted mean lengths of stay as maximum lengths of stay and interpreted case reimbursement levels as caps on per case expenditures. This practice is clearly driven by profit motives or confusion rather than by respect for the intention or the content of Medicare rules. Per-case payments based on a detailed analysis of what should be done for patients with a particular condition might also be interpreted as a Medicare cookbook for providing care. This criticism has already been raised about Young's (1985) design of normative weights for Patient Management Categories (PMCs) (Chapter 3.V).

In summary, a fully-specified system does not seem to be a realistic or desirable goal; PPS will continue to depend on averages.

IV. EVALUATING DRGs

In this section we provide a detailed evaluation of DRGs as a tool for predicting costs of care. In the next chapter, this evaluation of DRGs will also serve as a model for our evaluation of alternative systems (summarized in Table 2.1). Here, we describe how DRGs classify cases, the variants of DRGs which have been used, the face validity of the system, its data requirements, the publicly available documentation, problems of subjectivity and vulnerability to manipulation, the reliability of DRGs, and finally their ability to predict necessary care.

A. CLASSIFICATION STRUCTURE: NONORDINAL AND ORDINAL SYSTEMS.

Case mix systems can take two major forms.

1. Nonordinal Systems. DRGs, Disease Staging, and PMCs create a large number of categories which are clinically distinct. CSI creates such categories, but only as a route to a final severity score which is intended to be comparable access categories. Cases in each category are expected to require similar clinical management and resources but calibration with actual data is necessary to rank categories from most to least costly or severe.

2. Ordinal Systems. The APACHE score (APACHEII), Medical Severity Groups MEDISGRPS, SOII and CSI assign a single score to each patient, which measures how sick the patient is; Staging assigns a stage within the disease category. The score orders patients from least sick to most sick. APACHE, MEDISGRPS, SOII and CSI scores are generally used in conjunction with a disease oriented classification system.

Though an ordinal system which produces a single severity score appears to be very simple, the effective number of categories multiplies when the severity score is combined with diagnostic categories for payment purposes. DRGs, with 468 categories, are actually the most parsimonious of all the systems we shall examine. For payment purposes, the data available to determine payment rates for each category are the only limit to the number of categories.

Both ordinal and nonordinal systems (PMCs and SOII) may be unable to assign some patients; a nonordinal system may assign patients with multiple diagnoses to multiple categories (PMCs, Staging). If a system cannot assign a patient, there must be a fallback strategy for payment; if it makes multiple assignments, there must be a rule to determine payment for patients with multiple assignments. DRGs classify all patients and assign each patient to only one category.

The DRG system was developed at Yale University by John Thompson and

Robert Fetter. Health Systems International holds the current HCFA contract for updating the DRG classification program. The update system includes refinements of DRG definitions (e.g., HCFA recently changed the way in which patients can be classified in the kidney transplant DRG), creation of new DRGs (e.g., HCFA created a new DRG for multiple joint procedures), and restructuring of whole groups of DRGs (e.g., the entire group of substance abuse DRGs was recently restructured). Averill (1986) describes HCFA's current procedure for updating DRGs. DRGs are now used by Medicare and a number of other payors as a basis for payment and are the only classification system in use for payment.

The DRG system has a "tree" structure. The system classifies a case into one of 23 Major Diagnostic Categories (MDCs) on the basis of the principal diagnosis. Usually, MDCs are next divided into medical and surgical cases. Finally, medical and surgical cases are assigned to a DRG on the basis of presence or absence of certain procedures, age of the patient, specific principal diagnosis, presence or absence of a significant comorbidity or complication, and, in a few cases, whether the patient died, was discharged against medical advice, or was a transfer from another hospital. There are 471 DRGs, including DRG 468 for cases with surgery that is inconsistent with the principal diagnosis and several additional DRGs have been proposed. The DRGs were developed using clinical judgment in concert with statistical analysis of a large number of hospital discharges.

B. VARIANTS. All of the systems described in this report are still being developed, usually by the originator but sometimes by others as well. In consequence, a test of one version of the system may not apply to other versions, and a limitation in one version may not be present in another. Different versions usually constitute a developmental sequence, but in some systems (Staging, SOII) versions which are being developed along different lines are incompatible with one another. The existence of these variant versions complicates the evaluation of the case mix systems.

Due in part to its relatively long history and in part to its actual use in the New Jersey and Medicare PPSs, the DRG system exists in many versions; some, written for coding systems previous to ICD-9-CM (Commission on Professional Hospital Activities 1978), are primarily of historical interest. We will present only data from the ICD-9-CM grouper and will, when possible, distinguish among the sequence of revisions which HCFA has made since 1982.

C. FACE VALIDITY. Face validity is the apparent reasonableness of the data which the system uses to make classifications (to be discussed further in Chapter 3.VIII) and the way the data are combined to predict necessary care. There are five significant face validity issues for DRGs:

1. Data Limited to UHDDS. We noted that adopting DRGs limited PPS to using UHDDS data items and that these contain almost no direct evidence as to the severity of the illness and very little as to its acuity.

2. Using a Uniform List of Comorbidities/Complications for all MDCs. When a uniform list is used, a secondary diagnosis which is no more than a manifestation of the principal diagnosis may be interpreted as a comorbidity--for example, angina is a comorbidity if a patient has pneumonia but only a feature of the disease if the principal diagnosis is coronary atherosclerosis. This contrasts with the careful search for patterns of diagnosis which we will see in PMCs and Staging. Further, all comorbidities have equal weight--candidal otitis externa is given the same importance as candidal meningitis.

3. Use of Procedures for Classification. Using procedures for classification is, as noted above, a kind of cost-based reimbursement.

4. Categories Limited to About 500. As we have noted, this limitation restricts the ability of DRGs to deal with unusual conditions even when they can be clearly identified.

D. AVAILABILITY OF DOCUMENTATION. Systems differ in the degree to which they are in the public domain and the degree to which they are open for public inspection and research. There are several ways in which documentation can be unavailable:

-- A system may be copyrighted and the owners may refuse to allow any dissemination of the key materials (e.g., MEDISGRPS).

-- A system may be embodied in a complex computer program for which there is little documentation (e.g., PMCs).

-- A system may require special training of data coders for which no training manual is publicly available (e.g., SOII).

In each case, the ability of independent investigators to study the reliability or performance of the system is limited, and potential users are forced to rely for information on parties who have, at least, an emotional investment in the system and often a large personal financial stake in its adoption.

Extensive documentation about the DRG system is in the public domain. In addition to published articles (Fetter et al. 1980) and final reports of grant projects (Fetter et al. 1982), the PPS system itself is documented in the Federal Register (1983, 1984, 1985). Detailed descriptions of the workings of the DRG assignment algorithm and a computer program to implement the algorithm are available from a private company (Averill et al. 1983).

E. DATA REQUIREMENTS. All hospitals must provide the UHDDS (see Section I.A of this chapter) for each Medicare discharge and this information is available in computerized form. Some systems can classify patients using UHDDS data alone, others require data which are readily abstracted from the medical record, and others require data which can only be identified by specially-trained abstractors. A system using UHDDS imposes no data collection burdens and data to calculate payment weights for such a system are readily available. DRGs can operate completely from UHDDS.

Theoretically, the burdens of a system requiring special data collection could be reduced by collecting data on a sample of cases within each hospital rather than all cases, but the feasibility of this strategy cannot be estimated from available data for any of the systems which require special data collection.

F. SUBJECTIVITY, VULNERABILITY TO MANIPULATION, AND RELIABILITY.

Subjectivity and vulnerability to manipulation are not identical to unreliability. Unreliability means only that the system does not always make the same assignment for the same case; even abstracting a number, such as a patient's temperature, is subject to error and can lead to unreliability. But subjectivity means that the abstractor must make a judgment for which rules are not absolutely clear, so that subjectivity creates a risk of manipulation (conscious or unconscious) by abstractors with differing goals. Subjectivity can be discerned simply by examining the way the system is put together, while we must measure actual performance of the system to determine reliability.

1. Subjectivity. The danger of subjectivity is that a measure which is reliable in the hands of abstractors with shared goals may still be highly vulnerable to manipulation. In research, abstractors may consciously or unconsciously use the subjectivity of the system to make coding predict costs more accurately and thus to inflate the system's apparent performance. In payment policy, the important question is whether an abstractor who seeks to maximize Medicare payments (e.g., one working for the hospital) will code differently from one who seeks to minimize

Medicare payments (e.g., one working for Medicare). We must, therefore, be critically concerned as to whether there are procedures for resolving differences in judgment and whether they would hold up when large amounts of money were at stake.

DRGs require four judgments with subjective elements: making the diagnoses, coding the diagnoses, coding the procedures, and selecting the principal diagnosis.

-- Making diagnoses involves considerable clinical judgment and some kinds of judgment which are less clinical. Iezzoni and Moskowitz (1986) give the example of a patient with chest pain who might be coded as having chest pain, angina pectoris, or arteriosclerotic coronary disease, resulting in three very different payments. Although procedures for resolving diagnostic disagreement are better established than procedures for settling some other coding controversies, a subjective element remains. Demlo and Campbell (1981) found that independent coders could not resolve about 5-11 percent of disagreements, but the more explicit rules of ICD-9-CM may have reduced this ambiguity. One source of unresolvable ambiguity is that the diagnostician can often choose between a sign or symptom such as chest pain, coma, or anemia, and a pathological diagnosis such as myocardial infarction, drug overdose, or colon cancer (Iezzoni and Moskowitz 1986). Although Medicare rules defining principal diagnosis clearly give preference to listing the pathological diagnosis first, there is wide room for choice. Consistent "upcoding" of marginal cases can lead to higher case mix indices with the same mix of cases (Simborg 1981; Iezzoni and Moskowitz 1984, 1986, Gertman and Lowenstein 1984).

- Coding is also subjective because there is ambiguity, even in Medicare policy, as to how active a secondary diagnosis must be in order to be coded. A coding convention, which aggravates the problem, requires an abstractor to code any diagnosis which the physician says must still be ruled out at the time of discharge.

- Procedure coding involves significant subjectivity because DRGs require procedures coded in ICD-9-CM, and this procedure coding system is quite limited and somewhat obsolete. Thousands of dollars a case can ride on a controversy such as whether percutaneous transluminal coronary angioplasty is to be given the same code as coronary angiography (to which it is similar in method) or as coronary artery bypass grafting (to which it is similar in purpose). Such problems can be resolved definitively and subjectivity removed once they are identified and a rule is issued, but the system remains subjective in the interim.

- Sequencing diagnoses is also subjective and carries high cost implications because the first-listed or principal diagnosis is the major basis for determining DRG assignment. Coding conventions clearly require that the diagnosis found to be the cause of the admission is listed first, but the conventions are generally less clear on whether to list the underlying disease or the condition which caused the admission. For example, a patient admitted with pneumonia secondary to advanced cancer might be fairly given a principal diagnosis of either pneumonia or cancer.

2. Reliability. The reliability of a case mix classification system depends on the accuracy with which the necessary data are coded and abstracted from the medical record and the reproducibility of classification from the abstracted data. The reader should be alert to the way in which reliability is reported. If a system

has 95 percent type A cases (nonoutliers, for example) and 5 percent type B cases (outliers), two ratings could be 90 percent reliable and never agree on whether a case was type B. Since chance agreement is much less likely in a system with a large number of categories, some adjustment is needed to compare systems with different numbers of categories (for example, PMCs and SOII). The statistic which corrects for this error is called kappa, but it is not commonly used. Reliability may also be a more important issue for systems with a small number of categories which have very different payment implications than systems which have many categories with very similar payment implications (see discussion of coding accuracy below).

Computerized steps in classification tend to be reliable even if there are programming errors, although occasionally computer programs will give different results running on different computers. Experience indicates that programming errors affecting validity are relatively common, but those affecting reliability are rare.

Studies of DRG reliability (see Cretin and Worthman 1986) have focused on two questions: the accuracy of coding and changes in response to the incentives of PPS.

a. Coding Accuracy. Agreement in coding principal diagnoses using the older ICD-8 classification scheme is generally reported to be between 60 and 70 percent, depending on whether agreement is at the three digit or four-digit level (IOM 1977a; IOM 1977b; IOM 1980; Demlo and Campbell 1981). At the level of accuracy needed to assign a DRG, the agreement was just over 70 percent. It is important, however, to know the impact of disagreement on CMI. Using ICD-9-CM DRGs, Johnson and Appel (1984) found disagreement in DRG assignment in 47 percent of the cases, but the average impact on the CMI was 4 percent, with no more than a 2 percent difference in 77 percent of hospitals.

b. Response to Incentives. Carter and Ginsburg (1986) (Table 2.2) found that the CMI rose 9.2 percent from 1981 to 1984, of which they attributed 1.4 percent to a pre-PPS trend, 0.6 percent to PPS associated shifts to outpatient care, and 7.0 percent to coding changes. Of the 7.0 percent, a one-time jump of 4.0 percent resulted from shifting to a new abstracting form, which contained more information, but a further jump of 2.8 percent appears to reflect a self-limiting change in coding practices. Carter and Ginsburg (1986) could not determine whether this 2.8 percent change resulted from simply coding more carefully or whether it reflects "gaming," but 2.8 percent is less than the gap which Johnson and Appel (1984) found between actual coding and correct coding. There may also be room for further creep in the CMI because, even after the beginning of PPS, abstracts coded by hospitals for an abstracting service yield a 3 percent higher CMI than abstracts coded by the same hospitals for HCFA. Carter and Ginsburg point out that coding for Medicare was done sooner after discharge than coding for the abstracting service, and some Medicare coding is being done by hospital business offices rather than medical records personnel; either difference may explain the CMI difference.

G. PREDICTIVE POWER. In examining predictive power we will consider:

--what is being predicted,

--the data used,

--the statistical methods, and

--whether the prediction is for individual patients or for hospitals.

A system is more credible if investigators other than the developers have carried out studies and if separate data sets are used to create payment weights and to test predictive power. Predictive tests carried out on the same data set used in developing a system (even when that data set is "representative") usually make a system look better than it will look on a new set of cases. Although this type of bias can be reduced by the use of adjusted statistics, such adjustments are not always made.

1. What is Being Predicted. Although we would hope to evaluate DRGs against the gold standard of predicting necessary care, almost no evidence on this issue is available. Investigators have, at best, studied the ability of DRGs to predict actual costs and, in many cases, studies have examined only ability to predict charges or length of stay.

2. Data. The sample of patients and hospitals used to test performance can have two effects.

a. Exaggerated Apparent Performance. The performance of a system in classifying patients from just a few hospitals may give a biased indication of the performance of the system on a national sample of hospitals. This is especially true if all the hospitals in a sample are from the same region of the country or are about the same size or have the same level of teaching involvement. For example, if data come from hospitals with relatively homogeneous practice styles (compared to the whole country), the system may do much better than it would on a national data set. This may be why Horn found that DRGs accounted for 75 percent of cost variation in a sample of 15 "mostly large, teaching" hospitals, in contrast to the 25-40 percent of the variance more typically reported (Table 2.5).

b. "Volunteer" Effects. The performance of a system in a group of "volunteer" hospitals may not represent its potential in randomly selected or mandated hospitals. DRGs were developed on a sample from the nation's largest voluntary abstracting system; all other systems except PMCs were originally developed and tested on "volunteer" hospitals or (in the case of APACHE) on selected cases within "volunteer" hospitals. Often these "volunteers" have actually subscribed to the system for internal hospital management. These volunteers are likely to be an unrepresentative set of hospitals, and paying subscribers are likely to be much more motivated to make the system work accurately than hospitals seeking Medicare payments.

For these reasons, comparisons between two systems working on a single data set are much more informative than contrasting figures derived from different data sets. The variation in predictive power of DRGs between States (Table 2.3) suggests the risks of comparing performance of two different systems on different data. Because DRGs have been a benchmark for assessing performance of other systems, some of the results reported here were developed incidental to research on other systems.

We might expect that classification systems would perform better on systems where care was more uniformly appropriate. However, Kay et al. (1986) (Chapter 4.IV.B.1) found no improvement in the performance of DRGs in a data set from which apparently inappropriate days were removed compared to the same data set before inappropriate days were removed. This suggests that rigorously-defined appropriateness of care may not be necessary in data sets for comparing systems.

3. Statistical Methods. The reader needs to appreciate three features of the statistics used to measure system performance:

--whether they measure overall predictive power or measure ability to classify patients with patients whose needs are similar,

--whether they are highly sensitive to unusual cases, and

--whether they are applied to individual patients or to averages for hospitals.

a. Total Predictive Power v. Homogeneity of Categories. The two most frequently used statistics for measuring predictive power are

--explained variance (R^2), which measures the percent of total variation in cost which the system predicts, and

--coefficient of variation (CV), which measures the variation in cost among a group of patients as a percentage of the average cost for that group.

In general, R^2 comes closer to our intuitive notion of what a case mix measurement system should do, but CVs are frequently necessary when examining a part of a case mix measurement system or when the classification system cannot classify all patients.

b. Sensitivity to Unusual Cases. Both R^2 and CV are squared error measures. This means that variation from predicted cost (error) is squared before it is averaged. This practice is usual in statistics, but it means that a few unusual cases can dominate the results. For example, a system which predicts case costs exactly in 99 cases and misses by $1,001 in one case will appear to perform worse than a system which misses by $100 on all 100 cases. The impact of squared error methods is clear in Table 2.3, where removing small numbers of outliers dramatically improves the performance of case mix systems. The reader should be aware that the number of outliers removed in these "trimmed" studies is substantially smaller than the number which is paid under the HCFA outlier policy. Thus, in comparing methods the reader must be alert to whether the data have been trimmed in a uniform way. The reader should also be aware that using untrimmed data to compare DRGs to an alternative requires that DRGs classify outliers in a way which predicts cost. PPS, because of the outlier payment system, does not require that DRGs classify outliers in a way which predicts costs. Thus comparisons using untrimmed data are not ideal for deciding how classification systems will perform as part of PPS.

Another approach is to measure average deviation or average error, and methods which do this are often called "robust" because they are less sensitive to outlier cases. This approach is commonly used in economic impact studies (e.g., Horn et al's., (1985b) projection of the impact of using SOII to refine PPS and Mitchell et al.'s, (1985a) estimates of the impact of paying physicians with DRGs), and Cretin and Worthman (1986) use robust statistics to assess the impact of using Disease Staging to refine DRGs (Chapter 3). Cretin et al. (1986) use another kind of robust regression to assess the relationship of hospital characteristics to profit under PPS; here, outlier cases are given less weight in an otherwise conventional measurement. While robust statistics reflect the way hospitals compute profits and losses, they are not widely used or understood. In economics, however, computations often involve logarithms; these calculations tend to be more robust than calculations based on raw numbers.

c. Case Level v. Hospital Level Analyses. Because most experimental studies are conducted in just a few hospitals, the predictive power of the system is usually measured for individual patients. This approach gives more information about the ability of the system to protect individual patients against discrimination, skimming, and dumping. However, the aim of PPS is to pay hospitals accurately (Chapter 1), and hospital-level analyses are more revealing for this purpose (Chapter 4). In principle, one would expect the results to be quite different, with some kinds of variation averaging out at the hospital level and other kinds of variation being quite important between hospitals. For example, Medicaid eligible Medicare beneficiaries have almost exactly the same charges as other beneficiaries, but hospitals with a high percentage of Medicaid patients are reported to have higher costs per case (Gordon 1985, Altman 1985). Finally, patient level analysis is much easier to understand and assess: a clinician or even a layman can look at two patients and make some evaluation of whether they should have been paid equally. The same judgment is much more difficult when looking at whole hospitals.

4. DRG Performance at the Case Level. When viewed at the case level, using reduction in variance as the statistical criterion, DRGs are clearly helpful in explaining variations in resource use. In a variety of studies (Table 2.3) DRGs

explain from 17 to 30 percent of the variation in costs in untrimmed data and from 30 to 48 percent of the variation in cost on trimmed data (Mitchell et al. 1984; Mitchell et al. 1985a). Performance in explaining charges is similar to that for costs (West et al. 1985; Cotterill et al. 1985).

As we noted in section III.B.1.a, surgical DRGs consistently out perform medical DRGs (Table 2.4). At the case level, surgical DRGs explain 48 to 57 percent of the variance in charges in trimmed data, while medical DRGs only explain 9 to 16 percent of the variance (Mitchell et al., 1984, 1985a).

5. DRG Performance at the Hospital Level. The DRG-based hospital CMI accounts for 41 percent of variation in hospital average cost per Medicare case (Table 4.2). The CMI combined with the other variables used in reimbursement (the wage index, an urban/rural designation and the house staff/bed ratio), explains 66 to 69 percent of costs (Table 4.1) or charges (Cotterill et al. 1985) (Table 2.5).

One study of 15 hospitals reported that the DRG-based CMI alone explained 75 percent of the variance in cost per case (Horn et al. 1984b) (Table 2.5), but the sample was small and hospitals were more similar than in national samples.

H. SUMMARY. This chapter has examined the problems of predicting the cost of necessary care for individual patients. Section I laid out the predictive methods implicit in DRGs, Section II explored the ambiguities in the idea of necessary care and the policy problems latent in these ambiguities, Section III explored the issues in choosing variables to predict necessary care, and Section IV established a framework for comparing alternative case mix systems using the evaluation of DRGs as an example. In the next chapter we apply this framework to the evaluation of alternative case mix systems.

Table 2.1

BASIC FEATURES OF ALTERNATIVE CASE MIX MEASUREMENT SYSTEMS

SYSTEM NAME	CREATOR	ORGANIZATION	ORDINAL/ NON-ORDINAL*	NUMBER OF GROUPS/SCALE POINTS	DATA REQUIREMENTS	PROPRIETARY
Diagnosis-Related Groups (DRGs)	Thompson Fetter Averill	Yale	Nonordinal	468	UHDDS	No
Acute Physiology and Chronic Health Evaluation (APACHE)	Knaus Wagner	George Washington University	Ordinal	71	specially abstracted physiologic data	No
Disease Staging (Staging)	Gonella	Thomas Jefferson University	Nonordinal**	420 groups 4 levels each)	UHDDS	Software yes; architecture no
Medical Illness Severity Grouping System (MEDISGRPS)	Brewster	MediQual Inc.	Ordinal	5	specially abstracted history and physiologic data	Software and architecture secret.
Patient Management Categories (PMCs)	Young	Blue Cross of Western Pennsylvania	Nonordinal***	720	UHDDS	No
Severity of Illness Index (SOII)	Horn	Johns Hopkins University	Ordinal	5	specially coded scales "implicitly" integrated	No, but training manuals not public
Computerized Severity Index (CSI)	Horn	Johns Hopkins & Health Systems International	Nonordinal**	5	specially abstracted physiologic data	Incomplete

*In an ordinal system, patients have greater severity or care needs than patients with lower numbers and less than patients with higher numbers.
** Stages are ordinal within each staged disease category. CSI levels are ordinal within a diagnosis.
*** PMCs are ordinal within major disease category.

TABLE 2.2

DECOMPOSITION OF CMI INCREASE

In Percentage Points

CY 1981-FY 1984 CMI increase	9.2
Medical practice changes	
Pre-PPS trend	1.4
PPS-associated shifts to outpatient	0.6
Setting for lens procedures	0.7
Setting for chemotherapy	-0.1
Other outpatient substitution	0.1
Total, medical practice changes	2.0
Older patients	0.0
Coding practice changes	
PPS-induced	2.8
MEDPAR/PATBILL difference	4.0
Total, coding practice changes	7.0

NOTE: Numbers are multiplicative instead of additive.
For example, 1.028 * 1.040 = 1.070.

Source: Carter & Ginsburg 1986

TABLE 2.3

PERFORMANCE OF ICD-9-CM DRGs (CASE LEVEL)

		SAMPLE DESCRIPTION							
	SOURCE	SAMPLE	SOURCE	MEDICARE /ALL	TRIMMED /NOT	INDEPENDENT VARIABLE	DEPENDENT VARIABLE	R^2	AVERAGE COEFF VAR
1.	Coffey & Goldfarb 1984	394,000 Cases	Maryland 1979-81	Medicare	Untrimmed	420 DRGs	LOS	.16	94%
2.	Mitchell et al. 1984	272,087 Cases	New Jersey 1982	Medicare	Untrimmed Trimmed	DRGs	Costs	.18 .32	
		237,639 Cases	North Carolina	Medicare	Untrimmed Trimmed	DRGs	Costs	.16 .32	
3.	Mitchell et al. 1985	Not stated	Michigan 1982	Medicare	Untrimmed Trimmed	DRGs	Costs	.17 .30	
		Not stated	Washington 1982	Medicare	Untrimmed Trimmed	DRGs	Costs	.30 .48	
4.	West et al. 1985	28,047 Cases	South Carolina 1981	Medicare	Trimmed	DRGs	Charges	.26	
5.	Horn* et al. 1984a	19,122 Cases	1 University teaching hospital	All Patients	Not stated	436 DRGs	Charges	.33	85%
6.	Horn* 1985	5,580 Cases	1 University teaching hospital	All Patients	Not stated	401 DRGs	Charges	.29	76%
7.	Horn et al. 1984b	7,500 Cases	15 hospitals	All Patients	Untrimmed	DRG Cost Weights	Cost/ Case	.28	

*Published.

TABLE 2.4

MEDICAL AND SURGICAL DRGs

		SAMPLE DESCRIPTION								
	SOURCE	SOURCE	SAMPLE	MEDICARE /ALL	TRIMMED /NOT	DEPENDENT VARIABLE	MED	SURG	ALL	MEASURE OF VARIATION
1.	Mitchell et al. 1984	New Jersey 1982	272,087 Cases	Medicare	Trimmed	Charges	.11	.48	.32	R^2
					Untrimmed		.06	.31	.18	R^2
		North Carolina 1982	237,639 Cases	Medicare	Trimmed	Charges	.09	.51	.32	R^2
					Untrimmed		.04	.33	.16	R^2
2.	Mitchell et al. 1985	Michigan 1982	Not Stated	Medicare	Trimmed	Charges	.10	.49	.30	R^2
					Untrimmed		.06	.34	.17	R^2
		Washington 1982	Not Stated	Medicare	Trimmed	Charges	.16	.57	.48	R^2
					Untrimmed		.09	.39	.30	R^2
3.	West et al. 1985	South Carolina 1981	28,047 Cases	Medicare	Trimmed	Charges	.07	.55	.26	R^2
4.	Coffey 1985	Maryland 1979-81	268,358 Cases	Medicare	Trimmed	LOS	.12	--	--	R^2
							.859	--	--	Average Coefficient of Variation
			270,928 Cases	Medicare	Untrimmed	LOS	.09	--	--	R^2
							.955	--	--	Average Coefficient of Variation
5.	Frank & Lave 1985	HCFA	DRGs	Medicare		LOS	.945	.808	--	Average Coefficient of Variation
							(0.191)	(0.228)		SD
		HCFA		Medicare		Costs	.963	.801	--	Average Coefficient of Variation
							(0.194)	(0.703)		SD
		Maryland		Medicare	Untrimmed	Charges	0.971	.857	--	Average Coefficient of Avariation
							(0.666)	(0.292)	--	SD

TABLE 2.5

DRG CASE MIX INDEX (HOSPITAL LEVEL) WITH OTHER VARIABLES AS A PREDICTOR OF COST

	DATA				DEPENDENT	
SOURCE	SOURCE	SAMPLE	MODEL	COEFF	VARIABLE	PERFORMANCE
Pettengill & Vertrees 1982	MEDPAR 1979	351 DRGs	Case Mix Index (CMI)	1.081	Medicare Cost/Case	Adjusted $R^2 = 0.72$
			BLS wage index	1.000		
			Resident-bed ratio	.569		
			Bed size	.107		
			Large city	.149		
			Medium city	.037		
			Small city	.002		
Cotterill et al. 1985	MEDPAR 1981	358 DRGs	CMI (Cost weights)	2.891	Medicare Cost/Case	Adjusted $R^2 = .416$
			CMI (Cost weights)	1.012	Medicare Cost/Case	Adjusted $R^2 = 0.72$
			1981 BLS wage index	1.023		
			Resident-bed ratio	.580		
			Bed size	.119		
			Large city	.109		
			Medium city	.026		
			Small City	.001		
			Constant	7.322		
			CMI (Charge weights)	.969	Medicare Cost/Case	Adjusted $R^2 = 0.72$
			1981 BLS wage index	1.022		
			Resident-bed ratio	.545		
			Bed size	.117		
			Large city	.109		
			Medium city	.025		
			Small city	.000		
			Constant	7.334		
Horn et al. 1984b		15 hospitals	Case Mix Index		Hosptial Cost/Case	$R^2 = 0.75$
Thorpe et al. 1986	HCFA	3744 hospitals	Case Mix Index	1.246	Rand synthetic Costs	Adjusted $R^2 = 0.597$
			Resident/bed ratio	0.694		
			Wage Index	0.867		
			Urban/Rural	-0.141		
			Constant	7.690		

CHAPTER 3

EVALUATION OF ALTERNATIVE CASE MIX MEASURES

This chapter uses the framework established in the last chapter for examining DRGs to evaluate five case mix adjustment systems which have been proposed as replacements or refinements to DRGs--APACHE, Staging, MEDISGRPS, PMCs, and SOII. Section I describes three possible uses for alternative systems, Sections II-VI examine the five alternative systems, Section VII draws on this and the previous chapter to assess the value of different predictors of resource use, and Section VIII summarizes the usefulness of the systems for PPS.

I. CRITERIA FOR DIFFERENT SYSTEM USES

PPS needs case mix adjustment systems for research and for refining or replacing DRGs as payment tools; hospital managers need such systems for internal management. These uses place different demands on a system:

A. RESEARCH. PPS needs research instruments that can identify its strengths and shortcomings--criterion systems which are closer to the gold standard we have discussed. In some ways, a research system can meet less stringent requirements than a system for payment because it can:

--use data collection procedures which would be too cumbersome for a payment system, and

--use measures which are subject to manipulation (especially if the research data are collected retrospectively when the system cannot be "gamed").

Classification methods based on more subjective measures can be used in research, provided that care is taken in training and double checking the coders. But we have already observed that research coders may inadvertantly game systems to improve research results just as coders in a reimbursement system may game it. A research system clearly must be more valid or more reliable than DRGs to evaluate them.

B. REFINING AND REPLACING DRGs. A system for refining DRG payments need not be as precise as one for research because there is greater protection from the law of large numbers. A payment system must, however, be even more reliable and resistant to manipulation than a research system and it must produce appreciable improvements in the extent to which similar cases are classified together. A refinement system need not be able to classify all cases since unclassifiable cases can be paid under existing DRGs after appropriate recalibration, but a system to replace DRGs must be able to classify all cases.

C. MANAGEMENT. We have pointed out (Chapter 1.III.A) that the usefulness of a system to hospital managers is not a measure of its usefulness to PPS. Hospital managers need a system which is more precise than that used for PPS payment because they must assess performance of individual physicians and departments while PPS need only (Chapter 1) pay hospitals accurately. Further, hospital managers can use a system which is subjective and subject to gaming if their control of the abstracting process protects them from being gamed by physicians and departments. Finally, managers may be more interested in the system's value in quality assurance, a concern of greater interest to the PRO system than to the payment system. Thus, while we will examine use of the alternative systems for evaluation, refinement and replacement of DRGs for payment purposes, we will not focus on their usefulness in hospital management, and we will not consider the fact that some systems have proven useful to managers as evidence that they would be useful for PPS.

II. ACUTE PHYSIOLOGY AND CHRONIC HEALTH EVALUATION (APACHE)

Knaus and associates at George Washington University developed the APACHE system. APACHE was designed to assess the probability of death and severity of illness for Intensive Care Unit (ICU) patients (Wagner and Draper 1984; Knaus et al. 1981). APACHE has gone through several rounds of development and evaluation, and use in reimbursement is a recent proposal (Wagner and Draper, 1984).

A. CLASSIFICATION STRUCTURE. APACHE creates a score for the whole patient rather than classifying individual diseases or diagnoses, and the score is independent of diagnosis. APACHEII (the current version) builds a total score from three components:

1. The Acute Physiology Score (APS12), based on the values of 12 physiologic variables (e.g., temperature, pulse, respiratory rate, and arterial blood gases), which is determined within 24 hours of admission to the Intensive Care Unit (ICU). Each physiologic variable contributes points to the overall score according to a mathematical formula reflecting how far the patient's value deviates from "normal";
2. Age points for five categories of age; and
3. Chronic health evaluation points based on a history of organic insufficiency according to a list of conditions.

The overall score is intended to be an ordinal scale, which means that a patient with a higher score is more likely to die than a patient with a lower score, but that the difference is not necessarily proportional to the difference in score. APACHEII values can range from 0 to a theoretical maximum of 71; 55 is the highest value yet observed. The APS12 component can contribute 60 of the 71 points (Knaus et at. 1985, Wagner and Draper 1984).

B. VARIANTS. The original APACHE, with 34 variables in its physiologic component, has been replaced by the 12 variable APS score. The developers have suggested that a more parsimonious set of physiologic predictors could be developed, for reimbursement (Wagner and Draper 1984).

C. DATA REQUIREMENTS. APACHE requires physiologic and test data from the first 24 hours of admission to the ICU. Because these data elements are not routinely abstracted in most hospitals, customized abstraction forms are used to collect the data. Abstracting costs would be lower if some of the APS12 elements could be obtained directly from an automated laboratory reporting system or automated hospital record system. In the long run, as hospitals put more and more patient information on centrally accessible computer systems, a system like APACHE (or MEDISGRPS) might be redesigned to eliminate the need for special abstraction forms. Such technological changes could dramatically reduce costs. At the present, however, APACHE, like other systems using physiologic data, would create a significant abstracting and reporting burden.

D. FACE VALIDITY. Because APACHE was designed to assess the likelihood of death for ICU patients (Knaus et al. 1985), its origins are more explicitly clinical than those of any other case mix system, but this clinical origin may be less useful when assessing patients suffering from conditions with low mortality rates. APACHE's focus on critically ill patients explains the emphasis on physiologic measurements. The first version of APACHE was developed by clinicians who identified important variables and assigned relative weights to abnormal values. The initial physiologic score relied on 34 variables (Knaus et al. 1981). Subsequent analyses, using mortality as the predicted variable, reduced the number to 12 in APACHEII (Knaus et al. 1985).

Using APACHE to predict inpatient costs appears reasonable because of both the proportion of hospital costs devoted to critical care and the severity hypothesis discussed in the last chapter, but there are three problems.

1. Critical and Noncritical Care. The system was developed and tested on critical care cases, and APACHE's performance on non-ICU cases is unknown (the only study of this issue is Coulton et al. 1985). The relation of risk of death to

resource consumption may be very different in non-ICU patients.

2. Effects of Death. The confounding effects of mortality on resource use (Chapter 2.II.D) are an especially significant problem for a system designed to predict mortality. Wagner and Draper (1984) have shown that resource use actually declines in patients with the very highest APACHEII scores.

3. Goals of Treatment. In the group of patients to whom APACHE is especially directed, the clinical decision of whether to make the patient comfortable or to try to preserve life is critical to resource use (Chapter 2.II.C.3); the system cannot include this variable.

In sum, further research is necessary before we can be confident that APACHE will predict costs of care for the general patient population.

E. AVAILABILITY OF DOCUMENTATION. The APACHE system is reasonably well documented (Knaus et al. 1981, 1985), and has been implemented in many ICU settings (Wagner et al. 1984). The APS12 component of the score is particularly clear.

F. SUBJECTIVITY AND RELIABILITY. The variables used in the physiologic and age components of APACHE are quite objective, but the chronic health evaluation points require some judgment. Knaus et al. (1985) report 96 percent agreement on the 12 physiologic measures on repeat abstraction (Table 3.1), but "somewhat less" agreement on the Chronic Health Index; they did not report overall agreement between APACHE scores based on the two abstractions.

G. PREDICTIVE PERFORMANCE. APACHEII has been applied to ICU cases at 13 hospitals, mostly university teaching hospitals (Knaus et al., 1985; Wagner and Draper 1984). The original APACHE system has been applied more broadly (Knaus et al. 1981, 1984), including a recently reported international study involving 14 hospitals from the US, France, Spain and Finland (Wagner et al. 1984).

While the analysis of APACHE data has stayed in the hands of the developers, data collection has been carried out without their supervision (Knaus et al. 1985).

1. Performance at the Case Level. The main evidence on APACHE's predictive peformance is its ability to predict mortality (Table 3.2). As a predictor of death, the APACHEI score predicts 97 percent of deaths and 49 percent of those predicted to die do so (Knaus et al. 1981). In a study of 5030 patients in 13 ICUs, components of APACHEII, admitting diagnosis, and surgical status account for 32 percent of the variance in mortality, and 85 percent of predictions as to mortality are correct (Knaus et al. 1985).

Using the Therapeutic Intervention Scoring System (TISS) as a proxy for resource use, APACHEII with surgical status explained 17 percent of the variance in resource use for 5,790 patients in 13 ICUs (Wagner and Draper 1984). Only ICU cases were included and the patients were not limited to Medicare cases.

There are no studies of APACHE limited to Medicare patients or in non-ICU patients. Coulton et al. (1985) reported very uneven results when they applied APS12 in an ICU/non-ICU population; in one DRG the correlation between the APS12 score and resource use was negative.

2. Performance at the Hospital Level. Adding variables identifying different hospitals explained an additional 5 percent of the variance in TISS score in Wagner and Draper's study (1984), but they did not measure how much adding APACHE reduced variance in TISS score per admission across hospitals. Knaus et al. (1986) have used APACHE as an adjustor for studies of mortality across hospitals, but they did not test the ability of APACHE to account for interhospital variation in intensive care mortality rates. Their data did show, however, that observed

hospital mortality rates that were adjusted using APACHE varied much less than unadjusted mortality rates.

III. DISEASE STAGING (STAGING)

Gonnella and associates at Thomas Jefferson University developed the Disease Staging system; further refinements have been carried out by Conklin and associates at Systemetrics, Inc. Disease Staging, like APACHE, originated as a quality control system rather than a management or reimbursement system.

A. CLASSIFICATION STRUCTURE. Staging classifies diseases rather than patients and generally classifies them according to progression of an underlying disease rather than severity of acute presenting symptoms. Staging works in two steps:

--Each diagnosis is assigned to one of 420 Staged Disease Categories (SDC) according to its origins and pathology. SDCs are nonordinal; the category number says nothing about expected resource use.

--Within the SDC, the condition is then assigned to one of four stages of disease progression; stage four is always death. Stages within SDC are ordinal in that higher stages are always more clinically advanced.

If Staging is used for payment, a third step is necessary to classify patients who have more than one SDC.

The staging concept developed in clinical oncology, where stages represent progressive degrees of cancer spread and stage generally correlates with morbidity and mortality. Disease Staging generally preserves this meaning (Gonnella et al. 1984). The stages are:

--Stage I - conditions with no complications,

--Stage II - problems limited to an organ or system,

--Stage III - multiple site involvement,

--Stage IV - death.

For each disease, criteria for each stage are specified.

Because Staging criteria are disease specific, Stages are not directly comparable across diseases. Stage 2 hearing loss is very different from Stage 2 breast cancer (Gonnella et al. 1984; Conklin et al. 1984b).

Staging software stages all diagnoses. The system then offers two choices of the principal staged disease--either the staged principal diagnosis or the underlying staged disease; the latter is the highest staged SDC related to the principal diagnosis (based either on the principal diagnosis itself or on any secondary diagnosis related to the principal diagnosis). Information concerning staged conditions unrelated to the underlying staged disease may or may not be retained for further study (Louis et al. 1983, Conklin 1985).

In Staging, substages are often specified as well, reflecting clinical differences in progression. Substages generally follow the rule that higher numbers represent greater disease progression and a higher risk of morbidity and mortality (Gonnella el. 1983). Some substages, however, are not ordinal; they simply specify the locus of the disease more precisely.

B. VARIANTS. The two major variants are manual and computerized Staging. In addition, within the computerized version, there have been software modifications to embody alternative methods for handling principal diagnosis Staging, primary disease Staging, and the Staging of unrelated secondary diseases (Louis et al. 1983; Conklin et al. 1984a,b). Some ICD-9-CM codes had no corresponding SDC and could not be staged by earlier versions, but recent modifications of the system made possible staging of all diagnoses. These modifications also removed an earlier feature in which cancers had five stages.

The original manual Staging could be used to assign Stage at three points in a hospital stay: on admission, "peak" Stage, and at discharge. The computerized version generates a single assignment of Stage, which is probably closest to the "peak" stage (Arthur Young Co. 1983).

C. DATA REQUIREMENTS. The manual and computerized versions of Staging have different data requirements. The manual format provides detailed criteria for assigning stage and substage based on a review of the entire medical record. Manual Staging requires special training for coders.

The computerized version requires only UHDDS data to assign stages, although it can accept additional diagnoses and procedures.

D. FACE VALIDITY. Staging was developed with diverse goals, including utilization review, patient care evaluation, hospital management, and prediction of morbidity and mortality. The genesis of the system was entirely clinical, with panels of physicians developing both the definitions of the 420 Staged Diseases and the criteria for assigning Stage and Substage within each disease group (Gonnella et al. 1984). The diverse goals of Staging and the absence of resource data during its creation do not give Staging high face validity in predicting resource use. The Staging system ought to predict resource use when patients with more advanced disease require more resources, but sometimes the presence of acute, treatable clinical problems may predict need for care better than does the degree of disease progression. As with other case mix systems, conditions resulting in death will sometimes consume less resources than those with better outcomes; Stage 4 patients sometimes have lower costs than patients in Stage 3 (Conklin et al. 1984a; Conklin 1985).

The decision to minimize use of procedure codes is both a strength and a weakness. The strength is that the categories depend more on what is wrong with the patient and less on what is done, while the weakness is that a system of this kind seems likely (Chapter 2.II.A.1) to penalize referral surgical centers (Chapter 4.III reviews evidence that this would indeed happen). We will encounter this problem again with PMCs.

E. AVAILABILITY OF DOCUMENTATION. - Staging computer software and documentation are available for purchase. A complete set of manual Staging criteria are also available. Several researchers, besides the original developers of Disease Staging, have used either the manual (Arthur Young Co. 1983) or computerized versions (Coffey and Goldfarb 1984; Calore 1985).

F. SUBJECTIVITY AND RELIABILITY. Staging is vulnerable to the same subjectivity in coding diagnoses as is the DRG system (Chapter 2.IV.E). Although the system uses secondary diagnoses in a more precise way, incentives are still present to code additional secondaries and to "upgrade" coding in ambiguous situations. There are no studies documenting this risk.

Manual staging is more vulnerable to manipulation than the computerized version, because coders require special training and must review an entire chart. Preliminary evidence suggests that manual Staging is "moderately reliable" under research conditions (Richards et al. 1986).

Computerized Staging uses the same UHDDS data as does the DRG system, so inaccuracies of diagnosis coding will affect Staging; since Staging makes minimal use of procedures, inaccuracy in procedure coding is a smaller problem. We are unaware of studies of the impact of coding errors on a Staging CMI. When underlying staged disease is used, Staging is probably less vulnerable to manipulation of sequencing but more vulnerable to addition of marginal secondary diagnoses.

Information on the reliability of computerized Staging has not been reported, but Louis et al. (1984) have reported 78-80 percent agreement between manual and computerized Staging in assigning Stage (Table 3.1), depending on how one handles cases with incorrect prinicpal diagnosis.

G. PREDICTIVE PERFORMANCE. Data on predictive performance are limited to computerized staging. Several large data sets have been used by the developers of Disease Staging, including the National Center for Health Services Research Health Care Utilization Project (HCUP) file for 1977 (with data from 373 short-term non-Federal hospitals) (Gonnella et al. 1984). Other investigators have used Staging on large data sets: all Maryland Medicare discharges for 1979 through 1981 (Coffey and Goldfarb 1984; Coffey 1985), the HCUP file (Short and Coffey 1984) and a sample of 500,000 discharges from 24 teaching hospitals (Butler and Bentley 1982).

Medicare discharges from 54 Maryland hospitals for 1981 and from 80 New Jersey hospitals from 1979 have also been used in developing and testing a proposed modification of 17 DRGs using a Staging-based classification scheme (Conklin et al. 1984a, 1984b; Conklin 1985).

1. Performance at the Case Level. In both the Maryland Medicare data (Coffey and Goldfarb 1984) and in the Michigan data (Calore 1985), Staging performed less well overall than DRGs in explaining cost per case or length of stay (Table 3.3).

SysteMetrics developed alternative groupings which incorporated Stage and UHDDS variables (Conklin et al. 1984a; Conklin et al. 1984b; Conklin 1985) for 17 DRGs in 10 adjacent DRG clusters. (DRGs in an adjacent DRG cluster differ only in a single variable used to divide the DRG pair.) The alternative groups outperformed DRGs for cases in the original 17 DRGs (Table 3.4). These results suggest that Staging may be useful in refining highly variable DRGs.

2. Performance at the Hospital Level. All reported studies on reduction of variance and homogeneity have been at the case, rather than the hospital, level. However, Coffey and Goldfarb (1984) and Coffey et al. (1985) have examined the payment implications of using Staging rather than DRGs (see Chapter 4.IV). They

found that, at the hospital level, Staging would result in appreciable changes in payments compared to DRGs--apparently as a result of surgical cases. They did not, however, compare payments to actual costs.

IV. MEDICAL ILLNESS SEVERITY GROUPING SYSTEM (MEDISGRPS)

Alan Brewster and associates at St Vincent's Hospital, Worcester Massachusetts, developed MEDISGRPS, and Brewster is pursuing its development at MediQual, Inc. MEDISGRPS was designed for hospital quality control (Brewster et al. 1985b). MEDISGRPS is the newest and least tested of the systems described here.

A. CLASSIFICATION STRUCTURE. MEDISGRPS uses special abstracting forms to determine "key clinical findings" and a computerized algorithm to assign a patient to one of five ordinal severity levels based on the key findings. The levels are labeled 0 through 4, with higher numbers representing a higher probability of failure of one or more organ systems. The MEDISGRPS level, like that of APACHE and SOII, applies to the patient, rather than to the disease, and can be used to compare patients across diagnostic categories (Brewster et al. 1985b).

The abstraction form is designed to accommodate information on up to 500 different findings, including physical examination findings, laboratory, x-ray, and pathology results, and transfusions (Brewster et al. 1984a). The system generates a score based on the data which are available and does not require special tests or examinations of the patient.

MEDISGRPS is designed to use two reviews, one "on admission" (actually 48 hours after admission to allow relevant information to reach the patient's chart), and the second on the sixth post-operative day or the eighth day after admission. Data collected at the second review are also used to assess intrahospital morbidity or discharge status for those who have already been discharged (Brewster et al. 1985a). For patients with lengthy hospitalizations a third review may be carried out (Mediqual Systems, December 1984).

B. VARIANTS. MEDISGRPS has been revised a number of times, but differences between the versions have not been documented for the public. Subscribers receive the most recent version of the system, and analytic files have been updated to reflect this version. Apparently, changes involve the way computer software integrates data to produce a MEDISGRPS score rather than the raw data necessary to produce the score.

C. DATA REQUIREMENTS. The full version of MEDISGRPS requires two reviews of each patient's medical record (Brewster et al. 1985a; Brewster et al. 1985b), entry of data onto an abstraction form, and transfer of data from the form to a computer. Brewster et al. (1985a) estimate that abstraction and data entry require about 12 minutes per case. The cost of a single review for reimbursement has not been evaluated.

System costs include training abstracters, abstracting charts, and data entry for the computer. Given the large amount of input, the computational costs of this system may be higher than the computational costs of some of UHDDS-based systems such as DRGs or Staging.

D. FACE VALIDITY. MEDISGRPS, like all the other case mix methods reviewed here, used expert judgment in conceptualizing the system and defining the criteria for each classification level. Defining levels in terms of increasing probability of organ failure is appreciably different from describing either degree of disease progression or acute physical findings which predict the likelihood of death. The MEDISGRPS model avoids both the difficulties in using likelihood of death, which we reviewed regarding APACHE, and the problems with disease progression, which we reviewed regarding Staging. Although the comparability of levels across different kinds of disease is uncertain, this would not be critical if MEDISGRPS were used as refinement for DRGs or another disease-oriented system.

E. AVAILABILITY OF DOCUMENTATION. MEDISGRPS documentation is available only to individuals who sign an agreement not to reproduce or disseminate it. Full descriptions of the computer algorithm are not available even to subscribers to the system. This lack of documentation makes detailed assessment of the system impossible for outsiders. These difficulties may lessen when the system's proprietors are more certain how much they can reveal without compromising their commercial position.

F. SUBJECTIVITY, VULNERABILITY TO MANIPULATION, AND RELIABILITY. Two features of MEDISGRPS may make it vulnerable to manipulation:

-- The MEDISGRPS algorithm uses the most abnormal value when multiple values of a test are available (Brewster et al. 1985b), which may somewhat raise a patient's classification based solely on the number of tests done.

-- The system uses some data which are relatively subjective, such as physical findings and radiology interpretations. Such data may be stable in the absence of reimbursement incentives, but we have no data on whether payment incentives will change reports.

G. RELIABILITY. Information on the reliability of the MEDISGRPS abstraction process is sketchy. One report (Brewster et al. 1985b) notes that only 1 percent of the trainees failed to meet 95 percent accuracy standards after 9 weeks of experience (Table 3.1). Accuracy in this case was measured by comparing hospital trainee abstraction forms to those of an expert coder. As noted in the last chapter, reliability figures with small numbers of categories should be adjusted for chance agreement, which has not been done. Data on reliability of repeat ratings by the same rater are not available.

H. PREDICTIVE PERFORMANCE. MEDISGRPS was developed on a data set from one hospital (Brewster et al. 1985b). One subsequent report has included limited data from two additional hospitals (Brewster et al. 1984c).

1. Performance at the Case Level. MEDISGRPS has primarily been tested as a predictor of mortality (Table 3.5), morbidity (Table 3.6), and charges in selected conditions: abdominal pain, chest pain, shortness of breath and extremity pain. (Table 3.7 shows predictive power based on the first review and Table 3.8 shows predictive power based on the second review.) Total charges for different severity groups are significantly different.

Unpublished work by Moskowitz and Ash at Boston University using data from hospitals subscribing to the MEDISGRPS system, indicates that DRGs alone explain about 34 percent of the variance in charges while DRGs plus MEDISGRPS explain about 39 percent of the variance. MEDISGRPS has not yet been tested on a random or nationally representative sample of hospitals.

b. Performance at the Hospital Level. Reports on hospital level performance are not available.

V. PATIENT MANAGEMENT CATEGORIES (PMCs)

Wanda Young developed the system at Blue Cross of Western Pennsylvania specifically for cost analysis and hospital reimbursement (Young et al. 1982). PMCs are a nonordinal classification system accompanied by a normative but optional set of "resource weights."

A. CLASSIFICATION STRUCTURE. The PMC system is a clinically based system that assigns patients to one or more of 800 PMCs (Young 1984). PMC's were designed to create patient groups that are similar both in expected total resource use and in expected diagnostic and treatment services.

PMCs are grouped into 47 modules, each representing a disease or disorder. The first two digits of the PMC number refer to the module and are nonordinal. The second two digits identify a particular PMC within the module. The second two digits do not reflect a hierarchy, but the classification program is hierarchical in that the PMC to which a patient is assigned within a module is the PMC within a module which has the highest expected resource consumption of all the PMCs for which the patient may qualify (Young, 1985). The PMC classification software can assign a patient to several PMCs if multiple diagnoses lead to several modules. PMC assignments do not depend on the order in which diagnoses are listed, but the PMC associated with the principal diagnosis may be given priority for analytic or payment purposes. Procedures are used to make about 36 percent of PMC assignments; age and sex are used in a few instances.

The PMC system also specifies "components of care" for each PMC, a set of diagnostic and treatment services needed to provide effective care for an "average" patient. These components of care were constructed by about 50 panels of physicians and a normative "cost weight" is available for each PMC (Young et al., 1985). PMCs can also be used, like DRGs, with empirically-determined cost weights (e.g., Calore 1985). The method proposed by Young for handling cases with multiple PMC assignments combines the components of care for all assignments, deletes components of care which have been duplicated, and determines a cost weight from the remaining list of components.

B. VARIANTS. An early version of PMCs, which used reason for admission in addition to the usual UHDDS data, has been abandoned.

C. DATA REQUIREMENTS. PMCs operate on UHDDS data. If the alternative version is used, reason for admission must be hand abstracted.

D. FACE VALIDITY. The PMC system comprises a classification system and the normative "cost weights" based on components of care. The classification system could be valid even if the cost weights bore no relation to true resource needs, but the cost weights have no meaning outside the context of the classification system.

The PMC classification system was specifically developed to group patients according to their need for, rather than their use of, diagnostic and therapetic services. This clearly distinguishes PMCs from other systems reviewed here, which were designed to describe clinical severity, length of stay, or empirical costs. Against this more credible origin must be set the decision of developers of PMCs not to provide actual length of stay and resource data to their consulting clinicians as they were constructing the system.

The validity of the cost weights depends on the degree to which expert panels were able to reliably identify needed components of care and the appropriateness of the methods used to convert the specified components of care to costs. Since PMCs are the first comprehensive effort to assign normative costs to patient groups, we cannot estimate the degree to which different expert panels might arrive at similar estimates. As with construction of the classification algorithm, the fact that the panels worked without actual cost data means that they were not influenced by prevalent patterns of care. If prevalent patterns were appropriate, this decision probably put the developers at a disadvantage, but if prevalent patterns were inappropriate the decision would make PMCs more credible.

E. AVAILABILITY OF DOCUMENTATION. Because the PMC system was developed under a research grant and because the system was intended for use in reimbursement, both the computer program and descriptive documentation are available (Young et al. 1985). This documentation appears to be adequate for implementing PMCs. However, the source code for the algorithm is not available, and the program itself is not fully documented, so that an investigator cannot determine how most PMC assignments are actually made. This documentation deficiency poses significant problems for research and nonstandard uses of the system.

F. SUBJECTIVITY AND RELIABILITY. The reliability of PMCs, like that of other UHDDS-based systems, depends on the accuracy of the discharge data abstract. As with DRGs and computerized Staging, PMC's are vulnerable to upgrading of diagnosis codes; no estimate of the magnitude of this vulnerability is available. The fact that the system does not depend on the order of diagnoses decreases its vulnerability to manipulation of the sequence of diagnoses. On the other hand, this same feature may mean that the system is more influenced by listing marginal secondary diagnoses. Data on this issue are not available. Reabstracting reliability studies of the validity of PMCs are not available.

G. PREDICTIVE PERFORMANCE. PMCs have been applied to all discharges from 90 western Pennsylvania hospitals for 1981 through 1983 and to all discharges from 55 Maryland hospitals in 1983. PMCs have now been used both on the developmental data set (Young 1984, 1985) and on all Medicare cases from Michigan from 1982 (Calore 1985).

1. Performance at the Case Level. Young has tested the ability of PMCs to predict both mortality and length of stay (Table 3.9). The developers report statistics on the stability of death rates in selected PMCs over time (Young 1985) as evidence of the validity of the system. The primary performance analysis examined the homogeneity of length of stay within PMCs. The average coefficient of variation was 0.59 in the trimmed data set and 0.79 in the untrimmed data. This is

lower than the 0.94 for DRGs in untrimmed data sets reported by Coffey and Goldfarb (1984) and the 0.88 to 0.93 reported by Coffey and Goldfarb (1984) for Staging, but it is roughly comparable to the 0.82 mean coefficient of variation for DRGs found in untrimmed Medicare data (Cotterill et al., 1985).

Calore (1985) found PMCs slightly weaker than DRGs in explaining Medicare Part A (hospital) costs in 1982 Michigan data. PMCs explained 15 percent of the variance in cost per case in the untrimmed data set and 26 percent in the trimmed data set; DRGs accounted for 17 and 30 percent respectively. Calore did not use the normative PMC weights in this analysis.

2. Performance at the Hospital Level. There are no studies of the performance of PMCs at the hospital level.

VI. SEVERITY OF ILLNESS INDEX (SOII) and COMPUTERIZED SEVERITY INDEX (CSI)

Susan Horn at Johns Hopkins developed the SOII. SOII has been developed and promoted both as a case mix management system for hospitals (especially teaching hospitals) and as a possible refinement to the DRG system for use in PPS.

A. CLASSIFICATION STRUCTURE.

1. SOII classifies patients on a one to four ordinal scale which rates overall burden of illness rather than an individual disease (Horn et al. 1984a). A specially trained abstractor assigns the SOII based on a review of the whole medical record. The rater determines the SOII by "implicit integration" (Horn et al. 1983b) of seven severity dimensions, each of which also has four levels. The dimensions are:

-- Stage of the principal diagnosis (not the same as the Disease Staging assignment),

-- Complications of the principal diagnosis,

-- Concurrent interacting conditions affecting hospital course,

-- Dependency on hospital staff,

-- Extent of nonoperating room life support procedures,

-- Rate of response to therapy or rate of recovery, and

-- Residual impairment (Horn et al. 1985b).

In some applications, each level of the SOII is further divided into three parts: without an operating room procedure, with a moderate operating room procedure, and with a major operating room procedure (Horn et al. 1985b) using a list of ICD-9-CM procedure codes. The 12 categories of the procedure-adjusted SOII do not comprise an ordinal scale (Horn et al. 1984c).

Assigning the SOII requires a principal diagnosis (Horn and Horn 1985). In practice, the SOII is usually applied to patients after they have been grouped using a disease categorization scheme such as DRGs (Horn et al. 1985b; Horn 1985; Horn et al. 1985a; Horn and Sharkey 1983). Although SOII is generic and not disease-specific (Horn et al. 1985b), SOII levels may not be comparable across major disease categories.

2. CSI. Horn (1986) is now developing the CSI, which is so different from the SOII that it is grouped with it only for convenience. CSI divides all ICD-9-CM diagnoses into about 700 categories, and there is a specific set of severity criteria for each category. The criteria use clinical findings and the results of laboratory, X-ray, and electrocardiographic exams to assign a severity score to the principal diagnosis and all secondary diagnoses. A coder enters the diagnoses into a computer and the computer then requests the specific items which the coder must abstract from the chart. The computer then uses the criteria specific to each diagnosis to assign a severity level to each diagnosis. Finally, using a weighting algorithm, the program assigns an overall index to the patient. The criteria for individual

diagnoses, the weights assigned to the criteria in giving a severity score to an individual diagnosis, and the weighting scheme used for creating an overall severity index were all developed on clinical grounds but without interaction between actual data and the developers.

Horn has recently entered into cooperation with Health Systems International, a software firm, to produce the software for CSI, and field tests were scheduled to begin in October 1986. Whether the CSI would be used as a modification of DRGs or in conjunction with its own diagnosis categories is not yet clear.

In this discussion we will focus on SOII with some comment on how CSI differs from it.

B. VARIANTS. The published studies on SOII all use one or more versions of a manual scoring system. Within manual SOII, modifications have been almost continual since the system was first introduced. The major variant is the "procedure-adjusted SOII" described above.

C. DATA REQUIREMENTS. SOII differs from the other five systems both in the type of data input it requires and in the use of subjective human "integration" to arrive at scores. SOII requires that a specially trained abstractor review the medical record and record a rating on seven different four-point scales. The coding of these scales, unlike the coding of the data elements of APACHE or MEDISGRPS, depends not on locating specific clinical findings in the chart, but rather on the subjective integration by the abstractor of the entire content of the record. The developer reports that the abstraction takes a trained coder less than 3 minutes per chart (Horn 1985) when done as part of routine chart completion, which presumably means that uncomplicated cases can be coded very quickly. For a hospital with 2,000 Medicare cases per year, this represents about 100 additional hours of abstractor time per year. Other sources estimate that an average hospital would have to hire two additional abstractors (Nathanson 1985). Horn reports that abstractors can be trained quickly.

CSI will require manual abstraction of data and rater judgment as to whether medical record data satisfy the rating criteria; only the combination of criteria into codes and the combination of codes into a CSI rating is computerized. Thus, the level of computerization is similar to that of MEDISGRPS rather than that of DRGs, PMCs, or Staging. CSI is so different in concept and implementation from the manual SOII that it may be better to think of it as a separate system, rather than as a variation. The data collection burden of CSI relative to SOII is not yet defined.

D. FACE VALIDITY. Horn developed SOII as an alternative to DRGs for determining hospital reimbursement under the Maryland prospective payment plan (Horn 1985). Nevertheless, SOII was not created as a direct predictor of resource consumption but to measure the patient's burden of illness indirectly by asking "How sick is the patient?" (Horn et al. 1984b). If SOII measures this burden and if sicker patients need more resources, SOII can be expected to predict resource need. As we have already observed regarding Staging and APACHE, some very ill patients will have decreased resource needs, but this should not create greater problems for SOII than for other systems.

The seven subscales of SOII have face validity. Rate of response to therapy, interacting conditions, dependency and complications are all factors which should complicate treatment and increase costs. But the objectivity and measurability of these dimensions have less face credibility. The concepts are imprecise, it is difficult to define explicit criteria which are general enough to apply to all patients, and training manuals which are supposed to solve these problems are not available for public inspection.

The overall score is most often used in reporting data about the SOII, but the phrase "implicit integration" (Horn et al. 1983a) does not suggest that the overall scale is highly reliable. More recently Horn has described a mathematical formula for forming the overall score, but it is not clear whether this formula simply describes how rating usually works (Horn et al. 1985b) or prescribes how the overall score is to be given (Horn and Horn 1986).

A problem with the face validity of SOII is that response to treatment may reflect either the patient's illness or efficiency of treatment, while a major goal of PPS is to avoid paying for inefficiency. To a lesser extent, similar arguments apply to use of life support systems and extra nursing care. SOII presents a special problem because published data do not allow separation of scoring based on admission condition from scoring based on treatment. The SOII strategy, therefore, runs the risks which we reviewed in Chapter 2 when considering payment based on inhospital events. By contrast, CSI applies to a point in time and thus can be measured at admission, during the stay, or at discharge.

Until CSI is available for review, no independent judgment about its face validity or other characteristics is possible.

E. AVAILABILITY OF DOCUMENTATION. While much has been written about SOII, almost all of the reports to date have come from the original developer of the the index (Horn, et al., 1983b; Horn, et al., 1983a; Horn and Sharkey 1983; Horn, et al., 1984a; Horn et al. 1985a; Horn et al. 1985b). Because the system depends on the ability of the abstractors to reliably code the SOII score, the key to implementing the system is the training of the abstractors. Only limited documentation on abstractor training is publicly available, and no material is publicly available which would allow objective resolution of conflicts over SOII assignment.

The CSI is not yet documented.

F. SUBJECTIVITY, VULNERABILITY TO MANIPULATION, AND RELIABILITY. The SOII relies far more than other case mix methods on subjective and implicit judgments by the coder, a problem which is not, as we have pointed out, relieved by high reported reliability. Hospitals which have used SOII have been "volunteers" which adopted it for internal management. Medical records personnel have had strong incentives not to manipulate the system because they work for a hospital management which adopted the system in order to identify problems and discrepancies. This incentive would be absent if hospital management wished to obtain maximum reimbursement. The degree to which the system could be audited is unclear because of limited publicly available documentation. SOII has functioned to date by using Horn and her staff as the final arbiters of correct coding, but that solution appears impractical for a national payment system.

More information has been published on reliability of SOII than on any of the other systems being reviewed. The reported levels of agreement between expert coders and hospital based coders are high under Horn's supervision: between 88 and 93 percent (Table 3.1). These statistics are, however, uncorrected for chance agreement between raters. Although the frequency of agreement of SOII raters may be better (Horn and Horn 1986) than that for diagnostic coding, there is a difference between agreement when there are four choices (SOII) and when there are 10,000 choices (ICD-9-CM). Horn maintains that raters should be given credit for chance agreement in calculating reliability, but most other investigators do not agree.

The only reported use of SOII under supervision other than Horn's is the Graduate Medical Education Study (Arthur Young Co. 1983). In a preliminary analysis of a 10 percent sample of this data, Richards et al. (1986) found that 71.8 percent of SOII ratings were 1; 23.1 percent 2; 3.0 percent 3; and 2.2 percent 4. Of 577 cases rated more than once, 132 were rated twice, 417 three times, and 28 four times. There was complete agreement on 65.5 percent of cases, a disagreement of one point on 32.0 percent of cases, and a disagreement of more than 1 point on 2.6 percent. This corresponds to an interrater reliability of 78.9 percent, which is significantly lower than Horn has reported. Because of the very large fraction of cases rated 1, we would expect raters to agree by chance alone in 57.0 percent of cases, so the observed reliability means that raters agreed in only 50.9 percent of cases where they would not have done so by chance.

None of this analysis provides guidance on the most critical question, which is the impact of reliability issues on the payments which would be made if SOII were used for PPS. Such data are not available and, as Lloyd and Rissing (1985) have shown for DRGs, a relatively high error rate can have small impact on payment. Conversely, in a system such as SOII, where payment differences between SOII scores are large, the impact of moderate unreliability could also be large.

There is no information yet on the reliability of the CSI. Since it is based on adding a sixth digit to the ICD-9-CM codes, it presumably shares some of the same problems as DRG's, Staging, and other UHDDS systems as well as problems specific to CSI. On the other hand, the attention which CSI directs to specific physiologic criteria could lead to more accurate diagnoses.

G. PREDICTIVE PERFORMANCE. SOII research has been reported on a data set consisting of 70 percent of the discharges from 15 "primarily large teaching hospitals" (Horn et al. 1984b; Horn et al. 1985b). The abstractors were hospital employees who were trained under Horn's supervision, including regular follow-up reliability checks.

1. Performance at the Case Level. Severity procedure-adjusted DRGs explain 61 percent of the variance in resource use in the 15 hospitals (Table 3.11), while DRGs alone explained only 28 percent of the variance in the same sample. The

difference in performance between DRGs and SOII is probably somewhat smaller than it appears because the statistics appear not to have been adjusted for the smaller cell size in the SOII measurements. Severity-procedure-adjusted DRGs appear to give more weight to procedures actually performed than do current DRGs, but the use of operating room procedures to predict costs is already accepted in the DRG system.

No data are available yet on CSI and its performance, either at the case or hospital level.

2. Performance at the Hospital Level. In one study, Horn (1985b) converted charges to operating cost per case and carried out a hospital level analysis comparing DRGs to SOII-procedure-adjusted DRGs. In the 15 hospitals in this sample, SOII-procedure-adjusted DRGs explained 81 percent of the variance, which was not significantly better than the 75 percent of variance explained by DRGs alone. Horn reported that SOII-procedure-adjusted DRGs would come substantially closer to paying actual hospital costs than current DRGs but the methods for determining some of these payment implications were unclear.

VII. THE VALIDITY OF COST PREDICTORS

In this section we examine the accumulated evidence on the validity of certain predictors of need for care, independent of the particular system in which the predictors are imbedded. The validity of predictors is important both because it gives construct validity to a system constructed from these predictors and because it gives us guidance for future system construction. Valid predictors are also likely to lead to a clinically reasonable system, which is important for two reasons:

-- A clinically reasonable system is likely to retain its power even when relative costs of treatments change and even as clinical practices and technology evolve.

-- A clinically reasonable system is more likely to be acceptable to administrators and physicians and therefore less likely to stimulate gaming.

A. DIAGNOSIS. Principal diagnosis is not a strong predictor of use of services, as is clear from the weak performance of DRGs, Staging, and PMCs for nonsurgical discharges (Tables 2.4 and 3.12). In DRGs, which rely heavily on principal diagnosis, nonsurgical DRGs account for about 10 percent of variation in resource use. Staging uses patterns of diagnosis and performs slightly better than diagnoses as used in DRGs. PMCs, which use procedures somewhat less than do DRGs, do about as well as DRGs (Table 3.12). On the whole, evidence from these systems suggests that considering patterns of principal and secondary diagnoses may be slightly more powerful than the DRG strategy of resting assignment almost wholly on the principal diagnosis and relying on a uniform list of secondary diagnoses which will be considered comorbidities and complications. On the other hand, there is no evidence whether ignoring the order in which diagnoses are listed is a profitable strategy. This approach may decrease opportunities to manipulate the sequence of diagnoses while increasing opportunities to manipulate the system by adding inactive or marginal secondary diagnoses. Work in progress at Systemetrics will push the limits of combining Staging with DRGs. A recently funded cooperative agreement between HCFA and Robert Fetter at Yale proposes to develop lists of comorbidities and complications specific to each principal diagnosis, which might make DRGs more powerful.

One approach to the sequencing problem is to list the "primary" diagnosis first--that diagnosis which accounts for the greatest use of resources during the hospitalization. The Veterans' Administration has adopted this strategy, which specifically addresses the anomaly of a patient who develops a complication during the admission. The classic example is the patient who is admitted for cataract surgery and has a heart attack. But this strategy does not solve the other problems of sequencing or of diagnostic overlap.

To some extent the weak performance of diagnosis results from a diagnostic system (ICD-9-CM) which generally focuses on the pathology of disorder and its location rather than the severity of the disorder or its degree of progression (Worthman and Cretin 1986 and Averill 1986 discuss this limitation of ICD-9-CM in detail). But the weak performance of diagnosis also reflects the lack of information as to the purpose of admission.

B. SURGICAL PROCEDURES. Surgical procedures have considerable predictive power, for reasons described in the last chapter. In combination with diagnosis, they typically account for 30 percent of variation in resources used among all patients and 30-50 percent of variation among patients receiving surgery. Further use of procedures may further improve explanation of cost variation (Mitchell et al. 1985b, Cromwell et al. 1985), but this strategy risks an increasing return to cost-based reimbursement (Chapter 2).

C. SEVERITY OF ILLNESS ON ADMISSION. Evidence as to the explanatory power of severity of illness on admission is somewhat limited. Within certain groups (e.g., ICU patients), severity measures perform moderately well (Wagner and Draper 1984), accounting for 17 percent of the variance in resource use. On the other hand, Coulton et al. (1985) found very uneven results when applying physiologic measures within DRGs to a mixture of ICU and non-ICU patients. For certain conditions such as burns, multiple trauma, and low birthweight, the correlation of severity with cost of care is very clear for patients who survive. There is some evidence that admission severity measures do well across all patients (Brewster et al., 1985a,b).

The best publicized severity studies are of SOII (Horn et al., 1985a, 1985b), but these mingle severity on admission with severity during hospital stay so that the predictive power of admission severity cannot be estimated. Horn's methods are controversial (Davis 1985, Mullin 1985, Horn et al., 1985c). Nevertheless, there is now significantly more evidence than 3 years ago that clinical severity of illness on admission is strongly associated with overall costs of care. This same data, in addition to unpublished work by Moskowitz and Ash at Boston University, suggest that there are differences in average severity across hospitals which would be expected to lead to differences in cost. The work of Munoz (1985, 1986), which is discussed below, is consistent with this evidence. Generally, however, the variation in average admission severity across hospitals appears to be for smaller than the variation in admission severity at the individual case level.

Admission severity measures may become easier with time. As hospitals put more patient information on computer systems, a system like APACHE, CSI, or MEDISGRPS might be redesigned to eliminate the need for special abstraction forms and to abstract the necessary data from patient data files. Such technological changes in storage and retrieval of medical record information could dramatically reduce the marginal cost of implementing case mix adjustment methods using physiologic or other clinical data.

D. PSYCHOSOCIAL VARIABLES. Some authors have suggested that psychosocial variables such as poverty, living alone, being disabled, or being very old would increase a patient's need for care within a DRG (Stern and Epstein 1985). This is consistent with findings (Gordon 1985; Altman 1985) that hospitals with a disproportionate share of Medicaid patients have higher costs per admission than other hospitals with similar characteristics. To examine this question, Jencks and Kay (1986) examined the influence of a number of beneficiary characteristics on hospital charges, holding both hospital and case mix constant. They found that, in the most frequent DRGs, average charges for beneficiaries with Medicaid eligibility, previous hospitalization, age (80), extreme age (85), disability, and end-stage renal disease as well as beneficiaries admitted from a nursing home were all within 6 percent of charges for other beneficiaries in the same DRG (Table 3.14). While charges were higher for both nursing home residents and the aged, the two effects did not interact--charges for aged nursing home residents were no higher than those for other nursing home residents. While these findings do suggest that more careful adjustments for age may add some precision to DRGs, they do not support the hope that major improvements can be achieved by incorporating sociodemographic variables in the DRG system.

This comparison of patients within hospitals does not contradict findings that hospitals with a disproportionate share of Medicaid beneficiaries have higher costs than similar hospitals with fewer Medicaid beneficiaries and does not mean that the higher costs of disproportionate share hospitals are unwarranted. These findings do suggest that having a disproportionate share of Medicaid-eligible beneficiaries is a proxy for hospital and patient characteristics which result in higher costs rather than specific higher costs associated with Medicaid-eligible beneficiaries.

E. REFERRAL. Patients referred or transferred from other hospitals appear to have higher costs. Morrison et al., (1985) found that psychiatric patients received in transfer had charges about four times higher than direct hospital admissions. Work in progress (Bobula and Cotterill 1986) shows that hospitals receiving transfers have higher costs and that hospitals sending transfers have lower costs compared to the average hospital.

F. PURPOSE OF TREATMENT. Purpose of treatment is rarely even coded, and data linking purpose of treatment to costs are sparse. Garber et al., (1984), in a study of the difference between teaching and nonteaching services at a university hospital, found that, when patients were matched at admission, the mortality was lower and costs were higher on the teaching service. These data did not permit distinguishing quality of care from intent of care, but other evidence (including equal mortality at 9-month follow-up) suggested that the difference might be in therapeutic intent to keep dying patients alive.

G. KNOWLEDGE ABOUT THE PATIENT AT ADMISSION. Munoz et al., (1985, 1986) found higher costs for emergency room admissions in a sample of New York hospitals. He argued that need for inpatient diagnostic services in these patients contributed to increased costs, but this effect could not be separated from possible differences in severity of illness. Averill (1986) has suggested that diagnostic effort is an important determinant of cost after severity of illness has been considered, but there are no other studies to confirm that knowledge of the diagnosis at admission predicts cost.

H. IMPROVED CODING. An improved diagnostic coding system might better predict costs, since many of the case mix systems we have examined are really coding systems. ICD-9-CM was not designed to code either severity of illness or acuity of presenting symptoms, and a system which included either of these variables might work better. Horn's CSI system seeks to accomplish this goal. Authors who have criticized ICD-9-CM have suggested refining ICD on which it is based, but the next revision of ICD is due in 1995 and the lead time is so great that turning the revision toward measuring severity may not be possible even now. Furthermore, substantial changes in the ICD system might create serious conflict with epidemiologic uses of the system which require consistency from decade to decade.

Advocates of better coding have also pointed out that the procedure coding of

ICD-9-CM differentiates procedures by purpose rather than by what was done so that an endoscopic biopsy and a biopsy by operating room surgery can receive the same code. They suggest using a procedure code such as CPT-4 or HCPCS instead of ICD-9-CM. Nobrega et al., (1985) has shown that such a change is probably feasible, but there are no data yet to show that it would improve the predictive power of the system.

VIII. THE FEASIBILITY OF ALTERNATIVE CASE MIX SYSTEMS FOR PPS

In this section, we summarize the ability of alternative systems to replace DRGs, to refine DRGs, and to contribute to further research.

A. PERFORMANCE OF THE ALTERNATIVES AS DRG SUBSTITUTES. Examining the DRGs as case mix measurement tools raised concern as to their adequacy, but examining the alternatives underlines the substantial strengths of the DRG system. The basic criticism of DRGs is that they do not measure case mix well enough, but they are widely accepted as a "first cut," an acceptance illustrated by the degree to which they are taken as a point of departure for refinement systems.

By contrast, only two systems are plausible replacements for DRGs. SOII, MEDISGRPS and APACHE all require a diagnostic framework within which the severity level or score can be applied and could not by themselves replace DRGs. CSI aspires to be a potential replacement, but has not developed to a point where this potential can be evaluated.

Computerized Staging and PMCs are plausible substitutes for DRGs. Both are reasonably complete case mix adjustment systems, which have been developed and documented to the point that users other than the initial developers can implement the systems. Coffey and Goldfarb (1984) compared Staging to DRGs; Calore (1985) compared Staging and PMCs. Two findings emerge from these studies.

First, even though these two case mix methods are reasonably complete and documentation is available, they are not complete in the sense that DRGs are complete. The PMC software does not classify 5 (Young 1985) to 11 (Calore 1985) percent of cases. The PMC approach to patients with multiple patient management assignments requires accepting the structure of the normative PMC weights, and no data are available on performance of this approach. Staging now classifies almost all patients, but no mechanism is available to handle patients with multiple assignments and the risk of discrimination against referral surgical centers appears substantial.

The second finding is that neither Staging nor PMCs is better than DRGs (Table 3.10). This is particularly true of Staging, which actually explained about two-thirds of the variation in costs explained by DRGs. The difference in performance between DRGs and the other two systems clearly reflects the use of surgical procedures in assigning patients to DRGs: both Staging and PMCs perform about as well as DRGs when cases have already been classified as surgical or non-surgical (compare Tables 2.4 and 3.12).

Even though PMCs and Staging do not outperform DRGs, they might do better on specific tasks such as explaining the higher costs of tertiary care or teaching hospitals. Butler and Bentley, (1982) saw Staging as a system that would be better than DRGs for teaching and tertiary care hospitals (Butler and Bentley, 1982). However, when Coffey (1985) further analyzed the differences in the performances of Staging and DRGs, Staging did not explain the higher costs of teaching hospitals. Hospitals with a difficult surgical load also appear to lose under Staging, compared to DRGs. Pending further work that shows some new advantages of Staging, there does not seem to be a compelling reason to replace DRGs with Staging.

PMCs and Staging may have some structural advantages over DRGs for reimbursement because they generally do not rely on procedures (and therefore services rendered) for classification. It is especially promising that, using only UHDDS data, PMCs were able to perform as well as they did without a heavy reliance on procedures. In addition, the concept behind PMCs--grouping together patients on the basis of their needed components of care--comes closer to the goals of PPS than the approach used in DRGs--grouping patients based on actual resources used. Nevertheless, unless PMCs produce fairer payment than DRGs and do not discriminate against referral surgical centers, there is no compelling reason to put them in place of DRGs.

B. POTENTIAL OF THE ALTERNATIVE SYSTEMS FOR REFINING DRGs. All five systems might be useful in refining DRGs.

1. Staging. Systemetrics (Conklin et al. 1984a, 1984b; Conklin 1985; unpublished material) has developed refinements for about 60 DRGs based on Staging. These refinements rest on staging the principal and secondary diagnoses, assigning a weight to each secondary diagnosis based on regression analysis, and then forming subdivided DRGs. Although the computer programming is complex, the result is simply a somewhat larger number of DRGs in which comorbid and complicating diagnoses are used with greater sophistication. Preliminary results suggest that this approach increases the ability of the DRG system from about 36 percent of adjusted charges to about 41 percent of adjusted charges--an improvement roughly comparable to that achieved by MEDISGRPS but without need for special chart abstracting. Work on an earlier version of this refinement (Table 3.13; Cretin and Worthman 1986) suggests that this refinement would measure a somewhat more complex case mix in teaching hospitals and inner city hospitals. Because this refinement rests on use of secondary diagnosis while MEDISGRPS is

generic to all diagnoses, the two approaches should be complementary rather than directly competitive and might do well in combination. Some 6 to 12 months of additional work will be needed to confirm the initial findings and to define the interaction of Staging-based DRG refinements with MEDISGRPS.

2. SOII. The developer of the SOII has strongly advocated using SOII to modify DRGs. However, the subjective nature of the system and the absence of public training materials and formal methods to resolve conflict raise very serious obstacles to this strategy. Even the superficially objective DRG system suffered from an unexpected amount of "creep"; the more subjective SOII could well create more serious problems. The logistics of implementing and policing a manually coded, subjectively assessed national system would be extremely formidable. Horn's current focus on developing the more objective CSI suggests that she recognizes the importance of the subjectivity issue.

An equally serious problem is that SOII has been examined only in 15 primarily large, teaching hospitals, all of which had CMIs above 1.00 (Horn et al. 1985b). There is no evidence as to the redistributive effects of an SOII-based refinement among different classes of hospitals. When Disease Staging was tested only in teaching hospitals, it appeared to capture the range and severity of cases (Butler and Bentley 1982), yet when Staging was tested on statewide data, it did not appear to be an improvement over DRGs (Coffey and Goldfarb, 1984).

CSI is designed to remedy many of the obstacles to use of SOII, but assembling a national data base to test the proposed system will be both expensive and time-consuming. Speculating on the suitability of CSI as a refinement or replacment for DRGs is clearly premature.

SOII may be most valuable as a research and hospital management tool if technical questions regarding reliability measures can be resolved and if the risk of coding actual care instead of needed care can be shown to be controllable.

3. PMCs. PMCs might be used to refine DRGs somewhat as Staging has been used by Systemetrics. The fact that the two systems are very different in structure is less of an obstacle to such a strategy than the inadequate current state of PMC documentation. Katherine Calore at the Center for Health Economics Research has started to explore this approach.

4. APACHE. APACHE has demonstrated ability to predict costs in the critically ill, and it might be especially useful in paying for outlier cases, where data costs would be moderated by the small number of cases involved. This approach corresponds to concerns that hospitals might tend to avoid critically ill patients who have a high risk of becoming outliers. However, current data to assess this approach are not available.

5. MEDISGRPS. MEDISGRPS could improve the DRG system somewhat, but evidence that it can make payment more accurate must await collection of new data, such as that now being collected by the PROs for HCFA's Health Standards and Quality Bureau. The data burdens of MEDISGRPS for Medicare patients are not yet fully defined.

6. Conclusions Regarding DRG Refinements. Two types of refinements to DRGs can be considered: those which, like DRGs, only require UHDDS elements to operate and those which would require the expansion of UHDDS to include reason for admission, physiologic test results, or other elements. In the short run, UHDDS-based modifications are easiest to implement quickly.

The Staging-based modifications to DRGs are the most highly developed of available modifications. These refinements can be made using data which are currently collected but would require significant modifications to the Grouper software which assigns DRGs. They appear to appreciably increase the classification power of DRGs, but their impact across hospitals is uncertain. Evaluation of this modification using actual cost data from the first year of PPS appears to be desirable.

Other promising routes for DRG improvement include more precise adjustment for age, adjustments for transfers, inclusion of whether a patient came from a nursing home, and use of a better list of diagnoses for comorbidities and complications. The possibility that emergency room admissions are more expensive is intriguing but it is difficult to see how this could be used to determine payment without inviting manipulation.

Of the remaining systems (APACHE, MEDISGRPS, SOII and PMCs), APACHE seems closest to being implementable. MEDISGRPS (with its extensive data requirements) and the new CSI (with its requirement for implementing a new system of diagnosis and procedure coding) are not yet fully developed. SOII, with its labor intensive, subjective measures, may be useful in research and hospital management but appears unready for use in payment.

C. WORK IN PROGRESS AND FUTURE RESEARCH NEEDS. As other parts of this report suggest, a critical problem in DRG refinement is that we do know many problems in the current system but we do not know which are most important. If problems are concentrated in outliers (Chapter 4.VI), then systems with broad categories are likely to be relatively ineffective improvements. Results to date with Staging suggest that this is the case. If problems reflect differences in the purpose for which the patient is treated rather than the patient's condition, then new approaches to classification are needed and detailed attention to certain kinds of procedures are in order. If the difficulties result from differences in the degree of diagnostic and therapeutic aggressiveness between hospitals and between physicians, then Medicare will need either to pay for these differences without determining their efficacy or to find ways to pay for outcomes rather than treatments.

1. Work in Progress. A number of studies now in progress will yield significant clues as to the relative performance of a number of currently available measures. The Graduate Medical Education study will not only compare the performance of SOII and manual Staging (both now somewhat dated systems), but will look at the interplay between case mix, quality, cost, and appropriateness of care in a selected group of hospitals. Ashcraft and Thompson, at the University of Michigan, are pursuing case mix evaluation in a consortium of teaching hospitals. Alemi, at Tulane, is looking at the measurement of severity in myocardial infarction cases. Iezzoni, at Boston University, is examining case mix measures across hospital types. HCFA is sponsoring research at the Brandeis and Rand/UCLA Health Care Financing Research Centers to review the adequacy of the current case mix adjustment, looking for evidence of skimming and dumping, regional practice variations, and appropriateness of care.

These studies should provide significant guidance as to how much information physiologic variables can add to case classification and supply guidance for possible large-scale studies involving primary data collection.

2. DRG Refinement Using Staging and PMCs. Conklin (1985) has already demonstrated that Staging and DRGs can be synthesized to produce a system which is somewhat more powerful than current DRGs without requiring additional data collection and SysteMetrics is continuing work in this area. Young has viewed PMCs as a system so different from DRGs that synthesis is not possible, but PMCs may nevertheless provide useful tools in refining parts of the DRG system. Calore and colleagues at the Brandeis Center are exploring this approach. Although the improvements promised by these strategies appear to be modest, such

improvements also appear to be appropriate. Such steps appear to be more useful than comparisons of the existing DRG, Staging, and PMC systems on larger data sets since PMCs and Staging do not appear to provide sufficient advantages to consider using them as replacements for DRGs.

3. Sequencing Diagnoses and Improved Comorbidity/Complication Lists. A wide variety of evidence indicates that sequencing of diagnoses is inherently subject to ambiguity and error. Likewise, the lists of comorbidities and complications are oversimplified and nonspecific. Another source of work in this area is a cooperative agreement between HCFA and Robert Fetter at Yale University, one of the developers of the DRG system. This project also seeks to refine the use of comorbidities and complications in the DRG system, using an electic approach which draws on a number of sources. The award was made in August 1986, and the project is still in an early stage.

4. DRG Refinement Using Physiologic Measures. Studies in progress will provide considerable evidence on use of physiologic measures, but none will provide definitive information. This is due, in part, to the fact that the systems are evolving so rapidly so that new versions will almost certainly be in place by the time results from the current studies are analyzed. A second problem is that each system is developed and tested on a limited data set, so that cross-comparisons are difficult. Nevertheless, further information on whether incorporating physiologic measures into PPS is desirable.

-- A study on a nationally-representative group of hospitals and patients can give a picture of the payment impact of applying physiologic measures to PPS payment. There are insufficient data today to predict the payment impact of any of the systems which do not rest on UHDDS.

-- A collection of physiologic data and corresponding costs will be necessary in order to determine appropriate payments when physiologic measures are used for payment. There is currently no source for such "calibration data."

Designing a comparison requires approximate answers to three questions:

-- Should Medicare seek to adopt an existing "off the shelf" system or to construct a system which synthesizes elements of several competing designs? Given the problems with existing systems and the difficulties of publicly examining proprietary systems, this policy issue must loom large in the design of any comparison study.

-- For what patients will physiologic measures be most useful? Is it possible to focus on outliers or on patients with unusual conditions? A comparison study designed to pilot the adoption of a universal system will be very different in design from one intended to resolve payment problems for certain groups of patients.

-- What clinical data are most useful for classification? Can a system be confined to rigorously defined measures such as temperature and serum potassium, or must it use softer findings such as "pleural tightness" or "jugular-venous distension." The difficulties of using subjective elements are clear, but the feasibility of avoiding them is uncertain.

Studies now in progress by Alemi, Ashcraft, Iezzoni, and others will serve as pilot projects whose results will provide guidance on these three issues as well as other aspects of study design. The study itself will then allow HCFA to make an informed decision about incorporating physiologic measures into the payment system.

E. CASE MIX AND CAPITATION RATES. The problems of determining the correct payment for a hospital admission appear to have a natural relationship. Analysis to date suggests, however, that case mix refinement may have much less to contribute to capitation rates than one would hope. The problem is simply that the acute severity of illness and intensity of care which contribute so much to determining hospital costs are not useful for describing a beneficiary's expected use of services over the course of a year. Rather, capitated costs appear to be determined by chronic health status. For this reason, efforts to refine the measurement of comorbidities in the DRG system may be the most important common theme in refinement of DRGs and capitation rates.

F. CASE MIX MEASUREMENT AND MEASURING THE QUALITY OF CARE. The methods and objectives of case mix measurement are central to measuring the outcomes of care. The central feature of such measurements is to compare outcomes of care provided to similar patients by different providers or in different settings. The key problem is assuring that patients are indeed similar; measuring outcomes presents some difficulties, but certain important outcomes, such as death, can be ascertained with relative ease and great accuracy.

The best criterion for classifying patients for case mix measurement is the cost of efficiently providing necessary care, while the best criterion for quality measurement is the probability of some outcome (such as death or return to work or recurrent hospitalization or freedom from pain). Nevertheless, the problems are closely analogous because, as we have argued above, clinical homogeneity of case mix measurement groups is a natural and efficient basis for such groups. The emphasis on severity measurement in case mix measurement suggests how naturally clinical homogeneity has become an objective in case mix measurement.

Available measures are in many ways more applicable to quality of care studies than to case mix measurement at this time because studying the quality of outcomes or just a few conditions promises to be much more useful than improving case mix measurement for just a few conditions. This is true because there is enough cross-subsidization within hospitals to make the cost of treating a particular condition less informative about the overall efficiency of the hospital; the problem does not rise for quality of outcomes.

G. SUMMARY. The clearest conclusion of this chapter is that there is no persuasive alternative to DRGs and no revolutionary improvement is readily available. A series of modest improvements appear possible, and may, in combination, produce even larger improvements. Work to measure quality of outcomes is critically related to classification for payment because we can only design payment systems which maximize quality at minimum cost when we better understand what kinds of outcome are most affected by cost. Any project which significantly improves DRGs as a payment system is likely to yield substantial rewards for measuring quality of care.

TABLE 3.1

REPRODUCIBILITY

SOURCE	SAMPLE DESCRIPTION		WHAT IS COMPARED		MEASURE OF COMPARISON	RESULT
	SAMPLE	SOURCE	RATING (1) /	RATING (2)		
			PATIENT MANAGEMENT CATEGORIES			
Young . 1985	28,790 Cases	6 Hospitals in Western Pa. FY1981	PMCS Reason for Admission plus UHDDS	PMCS UHDDS only	Percent Agreement Case Classification	89.3%
			Criterion Assignment by Nurses	PMCS UHDDS only	Percent Agreement Case Classification	90% (After correcting ICD coding)
			Criterion Assignment by Nurses	PMCS Reason for Admission Plus UHDDS	Percent Agreement Case Classification	95% (After correcting ICD coding)
			APACHE II			
Knaus et al. 1985	5,030 Cases (Excludes coronary artery bypass surgery)	13 ICUs 1979-1982	Abstraction of phsyiologic items (APS12)	Second Abstraction physiologic items	Percent Agreement Item Abstraction	96%
			Abstraction of chronic health index	Second Abstraction Chronic health	Percent Agreement	"Somewhat less" than 96%
			SEVERITY OF ILLNESS INDEX			
Horn and Horn, 1985	1,775 Cases	18 hospitals	SOII rating by expert rater	SOII rating by hospital coders (Fourth check; >2 months experience)	Percent Disagreement Case Classification	7.4 percent

TABLE 3.1 (contd)

			SEVERITY OF ILLNESS INDEX (contd)			
Horn and Horn, 1985	4,489 Cases	18 hospitals	SOII rating by expert rater	SOII rating by hospital coders (Fourth and subsequent checks)	Percent Disagreement Case Classification	6.5% Weighted Average
			MEDISGRPS			
Brewster et al., 1985b	200 coders.		MEDISGRPS abstract by expert abstractor	MEDISGRPS abstract by hospital trainee	Percent trainees failing to meet 95% accuracy after 9 weeks experience	1%
			abstract by hospital abstractor	abstract by second hospital abstractor	Contractual standard for agreement	95%
			DISEASE STAGING			
Louis et al. 1984	1,380 Cases	5 hospitals	Stage by manual abstraction	Stage by computer on UHDDS data	Percent agreement on stage (Includes effect of cases with incorrect principal diagnosis)	78% weighted
					Percent agreement on stage (exludes effect of incorrect principal)	80.5% weighted

Source: Cretin et al. 1986

TABLE 3.2

PERFORMANCE OF APACHE II
(CASE LEVEL)

SOURCE	SAMPLE DESCRIPTION			INDEPENDENT VARIABLE	DEPENDENT VARIABLE	PERFORMANCE
	SAMPLE	SOURCE	MEDICARE /ALL			
Knaus et al. 1985	5,030 Cases*	13 ICUs 1979-1982	All payers	APS12 Age Chronic health Surgical stat ICU Admit dx	Mortality	$R^2=.32$ (85.5%) Cases Correctly Classified
Wagner and Draper 1984	5,790 Cases	13 ICUs 1979-1982	All payers	APS12 Age Chronic health Surgical stat ICU Admit dx Hospital	"Cost" (Trimmed TISS**)	$R^2=.22$
				APS12 Age Chronic health Surgical stat	"Cost" (Trimmed TISS**)	$R^2=.17$

---

* Sample excludes Coronary Artery Bypass Surgery.

** Therapeutic Intervention Scoring System (TISS). A value of 350 was assigned to TISS scores greater than 350.

Source: Cretin et al. 1986

TABLE 3.3

PERFORMANCE OF DISEASE STAGING
(CASE LEVEL)

SOURCE	SAMPLE DESCRIPTION				INDEPENDENT VARIABLE	DEPENDENT VARIABLE	R SQUARED	AVERAGE COEFF VAR
	SAMPLE	SOURCE	MEDICARE /ALL	TRIMMED /NOT				
Coffey & Goldfarb 1984	394,000 Cases	Maryland 1979-81	Medicare	Untrimmed	Stages (Principal Diagnosis), 805 groups	LOS	.12	88%
					Stages 698 groups	LOS	.10	93%
Calore 1985	300,122 Cases	Michigan 1982	Medicare	Untrimmed	Stages	Costs	.10	
				Trimmed			.17	

Source: Cretin et al. 1986

TABLE 3.4

Comparison of Percent Variance Explained (R Squared) and Percent Absolute Deviation Explained for 10 DRGs and Systemetrics Adjacent DRGs
Dependent variable: Synthetic costs/case

Case level Analysis

	R-squared Statistics (% variance explained)		Sum of Absolute Deviation Statistics (% Abs. Dev. explained)	
	DRGs	Systemetrics	DRGs	Systemetrics
Major Rec.Vasc. (DRG 110 & 111)	.0195	.1352	.0117	.0838
Biliary Tract (DRG 197 & 198)	.0333	.0847	.0345	.0792
Diabetes (DRG 294 & 295)	.0000	.0713	.0000	.0414
Major Bowel Proc. (DRG 148 & 149)	.0156	.0701	.0171	.0479
Kidney and UTI (DRG 320,321,322)	.0100	.0455	.0074	.0339
Benign Prostate Dis. (DRG 348 & 349)	.0109	.0398	.0058	.0223
Major Chest/Lung (DRG 75)	--	.0242	--	.0187
Cirrhosis & Alcohol. (DRG 202)	--	.0214	--	.0191
Cereb. vasc. ex TIA (DRG 14)	--	.0191	--	.0126
Pulmonary Embolus (DRG 78)	--	.0141	--	.0061

Sample: 20 percent sample of Medicare cases in FY 1984
Source: Cretin et al. 1986

TABLE 3.5

PERFORMANCE OF MEDISGRPS IN PREDICTING MORTALITY (CASE LEVEL, FIRST REVIEW)

SOURCE	SAMPLE DESCRIPTION				INDEPENDENT VARIABLE	DEPENDENT VARIABLE	MANN-WHITNEY U STATISTIC
	SAMPLE	SOURCE	MEDICARE /ALL	TRIMMED /NOT			
Brewster et al. 1985b	1,467 Cases (Abd pain)	Saint Vincent Hospital FY1983	All payers	Excludes Transfers	MEDISGRPS (5 groups)	Mortality n=43	Value = 13327 p = <0.01
	1,246 Cases (Chest pain)					n=58	Value = 10311 p = <0.01
Brewster et al. 1984a	1,313 Cases (Shortness of breath)	Saint Vincent Hospital FY1983	All payers	Excludes Transfers	MEDISGRPS (5 groups)	Mortality n=130	Value = 42432 p = <0.01
Brewster et al. Undated mimeo	1007 Cases (Extremity pain)	Saint Vincent Hospital FY1983	All payers	Excludes Transfers	MEDISGRPS (5 groups)	Mortality n= 24	Chi Square p = <0.01

Source: Cretin et al. 1986

TABLE 3.6

PERFORMANCE OF MEDISGRPS IN PREDICTING MORBIDITY* (CASE LEVEL, FIRST REVIEW)

SOURCE	SAMPLE DESCRIPTION				INDEPENDENT VARIABLE	DEPENDENT VARIABLE	MANN-WHITNEY U STATISTIC
	SAMPLE	SOURCE	MEDICARE /ALL	TRIMMED /NOT			
Brewster et al. 1985b	1,424 Cases (Abd pain)	Saint Vincent Hospital FY1983	All payers	Excludes Deaths and Transfers	MEDISGRPS (5 groups)	Morbidity n=204	Value = 70060 p = <0.01
	1,188 Cases (Chest pain)					n=156	Value = 40294 p = <0.01
Brewster et al. 1884a	1,183 Cases (Shortness of breath)	Saint Vincent Hospital FY1983	All payers	Excludes Deaths and Transfers	MEDISGRPS (5 groups)	Morbidity n=293	Value = 74450 p = <0.01
Brewster et al. Undated Mimeo	970 Cases (Extremity pain)	Saint Vincent Hospital FY1983	All payers	Excludes Deaths and Transfers	MEDISGRPS (5 groups)	Morbidity n=101	Chi Square p = <0.01

*This a system designed to use 2 reviews. Morbidity refers to failure to recover by second review.

Source: Cretin et al.

TABLE 3.7

PERFORMANCE OF MEDISGRPS IN PREDICTING CHARGES
(CASE LEVEL, FIRST REVIEW)

SOURCE	SAMPLE DESCRIPTION				INDEPENDENT VARIABLE	DEPENDENT VARIABLE	R SQUARED	AVERAGE COEFF VAR
	SAMPLE	SOURCE	MEDICARE /ALL	TRIMMED /NOT				
Brewster et al. 1984b	1,467 Cases (Abd pain)	Saint Vincent Hospital FY1983	All payers	Excludes Transfers	MEDISGRPS (5 groups)	Charges	.23	1.46 Total 1.11 Weighted ave for subgroups
	1,246 Cases (Chest pain)						.14	1.11 Total .80 Weighted ave for subgroups
Brewster et al. 1984a	1,183 Cases (Shortness of breath)	Saint Vincent Hospital FY1983	All payers	Excludes Deaths and Transfers	MEDISGRPS (5 groups)	Charges	.10	1.23 Total 1.03 Weighted ave for subgroups
Brewster et al. Undated mimeo	983 Cases (Extremity pain)	Saint Vincent Hospital FY1983	All payers	Excludes Deaths and Transfers	MEDISGRPS (5 groups)	Charges	.18	1.17 Total 1.05 Weighted ave for subgroups

Source: Cretin et al. 1986

TABLE 3.8

PERFORMANCE OF MEDISGRPS IN PREDICTING CHARGES
(CASE LEVEL, SECOND REVIEW)

SOURCE	SAMPLE DESCRIPTION				INDEPENDENT VARIABLE	DEPENDENT VARIABLE	R SQUARED	COEFF VAR
	SAMPLE	SOURCE	MEDICARE /ALL	TRIMMED /NOT				
Brewster et al. 1985b	1,424 Cases (Abd pain) 204 with morbidity at second review	Saint Vincent Hospital FY1983	All payers	Excludes Deaths and Transfers	MEDISGRPS (5 groups)	Charges	.41	1.46 Total .85 Weighted ave for subgroups
Brewster et al. 1984a	1,188 Cases (Chest pain) 156 with morbidity at second review	Saint Vincent Hospital FY1983	All payers	Excludes Deaths and Transfers	MEDISGRPS (5 groups)	Charges	.28	1.11 Total .67 Weighted ave for subgroups
	1,183 Cases (Shortness of breath) 293 with morbidity at second review						.27	1.23 Total .81 Weighted ave for subgroups
Brewster et al. Undated mimeo	983 Cases (Extremity pain) 102 with morbidity at second review	Saint Vincent Hospital FY1983	All payers	Excludes Deaths and Transfers	MEDISGRPS (5 groups)	Charges	.33	1.17 Total .97 Weighted ave for subgroups

Source: Cretin et al. 1986

TABLE 3.9

PERFORMANCE OF PATIENT MANAGEMENT CATEGORIES
(CASE LEVEL)

SOURCE	SAMPLE DESCRIPTION				INDEPENDENT VARIABLE	DEPENDENT VARIABLE	R SQUARED	AVERAGE COEFF VAR
	SAMPLE	SOURCE	MEDICARE /ALL	TRIMMED /NOT				
Calore 1985	300,122 Cases	Michigan 1982	Medicare	Untrimmed	PMCs	Costs	.15	
				Trimmed			.26	
Young 1985	489,721 Cases	Western Pennsylvania 1983	All payers	Untrimmed	Single Assignment PMCs	LOS		.79 Unweighted .76 Weighted by pts
				Trimmed				.59 Unweighted .56 Weighted by pts

Source: Cretin et al. 1986

Table 3.10

COMPARISON OF THE VARIATION (R-SQUARE) IN COST OF INDIVIDUAL CASES EXPLAINED BY THREE CASE MIX MEASURES

	Diagnosis-Related Groups	Patient Management Categories	Disease Staging
Trimmed	0.30	0.26	0.17
Untrimmed	0.17	0.15	0.10

* Cost data are untrimmed and exclude DRGs 468 and 470.

Source: Calore, unpublished data 1985.

TABLE 3.11

PERFORMANCE OF SEVERITY OF ILLNESS INDEX (SOII)
(CASE LEVEL)

SOURCE	SAMPLE DESCRIPTION				INDEPENDENT VARIABLE	DEPENDENT VARIABLE	R SQUARED	AVERAGE COEFF VAR
	SAMPLE	SOURCE	MEDICARE /ALL	TRIMMED /NOT				
Horn et al. 1984a	19,122 Cases	1 University teaching hospital	All Patients	Not stated	SOII and Procedure type (12 groups)	Charges	.49	70%
Horn 1985	5,580 Cases	1 University teaching hospital	All Patients	Not stated	SOII and Procedure type	Charges	.46	64%
Horn et al. 1984b	7,500 Cases	15 hospitals	All Patients	Untrimmed	Severity/proc adjusted DRG cost weights	Cost/ Case	.61	

Source: Cretin et al. 1986

TABLE 3.12

MEDICAL AND SURGICAL PMCs, DISEASE STAGING AND DRGs

	SAMPLE DESCRIPTION								
SOURCE	SOURCE	SAMPLE	MEDICARE /ALL	TRIMMED /NOT	DEPENDENT VARIABLE	MED	SURG	ALL	MEASURE OF VARIATION
				PATIENT MANAGEMENT CATEGORIES					
Calore 1985	Michigan 1982	300,122 Cases	Medicare	Trimmed	Charges	.11	.51	.26	R Squared
				Untrimmed		.07	.35	.15	R Squared
				DISEASE STAGING					
Calore 1985	Michigan 1982	300,122 Cases	Medicare	Trimmed	Charges	.13	.49	.17	R Squared
				Untrimmed		.09	.35	.10	R Squared
Coffey 1985	Maryland 1979-81	268,358 Cases	Medicare	Trimmed	LOS	.09	--	--	R Squared
						.847	--	--	Average Coefficient of Variation
		270,928 Cases	Medicare	Untrimmed	LOS	.07	--	--	R Squared
						.916	--	--	Average Coefficient of Variation
				DRGs					
Calore 1985	Michigan 1982	300,122 Cases	Medicare	Trimmed	Charges	.10	.49	.30	R Squared
				Untrimmed		.06	.35	.17	R Squared
Coffey 1985	Maryland 1979-81	268,358 Cases	Medicare	Trimmed	LOS	.12	--	--	R Squared
						.859	--	--	Average Coefficient of Variation
		270,928 Cases	Medicare	Untrimmed	LOS	.09	--	--	R Squared
						.955	--	--	Average Coefficient of Variation

TABLE 3.13

EFFECT ON HOSPITAL PAYMENT OF USING SYSTEMETRICS REFINEMENTS INSTEAD OF DRGs
For Seventeen DRGs, by Hospital Characteristics

Hospital Characteristic	Change in Payment Per Case ($)
Ownership (Omitted class: Voluntary)	
Proprietary	18.00
Public	15.00
City Size (Omitted class: Rural)	
Small City	0.00
Medium City	-10.00
Large City	-3.00
PPS 2nd quarter	24.00*
PPS 3rd quarter	-14.00
PPS 4th quarter	-30.00*
Trauma center	-56.00*
Number of beds (change per bed)	0.07**
Resident/bed ratio	341.00**
Proportion of Medicaid Admissions	98.00
Inner City	45.00*
Interaction of Prop. of Medicaid Admissions and Inner City	-216.00
New England	92.00**
Mid Atlantic	52.00
Florida	29.00
Illinois/Michigan	14.00
Indiana/Ohio/Wiscon.	-9.00
Mississippi	-30.00
Kentucky/Tenn./Alabama	-12.00
Neb./N. Dak./S. Dak.	7.00
Iowa/Kansas/Minn./Missouri	46.00*
Texas	43.00*
Ark./La./Oklahoma	4.00
Utah	109.00*
Ariz./N. Mex./Nev.	50.00
Idaho/Montana/Colorado/Wyom.	34.00
California	21.00
Alaska/Hawaii	407.00**
Oregon/Washington	78.00**
Estimated data1	15.00
Estimated data2	71.00**
Intercept	-68.00**
R Square	0.054

*P<.05.
**P<.01

Note: Results based on provider level models weighted by number of bills. See Table 3.4 for a list of ADRGs included.
Source: Cretin et al. 1986

Table 3.14

RELATIVE CHARGES FOR BENEFICIARIES WITH EIGHT CHARACTERISTICS
Compared to Otherwise Similar Beneficiaries
Controlling for DRG and Hospital in which Care was Rendered
5 Percent Sample of 1983 Medicare Discharges

Beneficiary Characteristic	Relative Charge
1. Nursing Home Resident.	1.062***
2. Hospitalization in Previous Year.	1.033***
3. Medicaid Enrollee	0.985**
4. Aged (>79).	1.038***
5. Very aged (>84).	0.987*
6. Younger disabled <65)	0.993
7. Younger endstage renal disease.	1.052**
8. Older disabled. (>64)	1.067***

* = p <.05 that ratio is = 1.00; ** = p <.01 that ratio is = 1.00;
*** = p <.001 that ratio is = 1.00, using t test.

[1] Based on 10 random samples of 60 DRGs each.

Source: Jencks & Kay 1986.

CHAPTER 4

HOW DO COSTS OF TREATMENT DIFFER BETWEEN HOSPITALS AND WHY?

The last two chapters examined case mix adjustment at the individual case level. We concluded that some increase in precision of case classification is possible, but that no system is currently available to either replace DRGs or to provide a strong severity adjustment to the DRG system. We suggested that work already in progress promises to improve case mix measurement as it applies to nonoutlier cases and that severity measures may do more for our understanding of outlier cases than of typical cases. But we also pointed out that the severity model must be supplemented by an understanding of the issues of intensity of care if we are to understand differences in cost per case.

Chapter 1 pointed out that hospital-level measures are an important complement to case mix adjustment at the individual case level. In this chapter we examine ways in which hospital-level adjustments may be able to achieve refinements which are impractical at the individual case level. Section I describes the hospital-level adjustments which were included in PPS and why they are constructed as they are; Section II provides a model of why hospitals have different case mixes; Section III examines the empirical evidence on hospital differences; Section IV looks at the empirical problems involved in setting prices which will be used across hospitals and how these issues influence our ability to compare hospitals; Section V analyzes information on the winning and losing hospitals during the first year of PPS; Section VI provides a parallel examination of the impact of the outlier payment system in the first year of PPS; Section VII discusses new ways in which hospital-level adjustments might be made; Section VIII examines the evidence that larger hospitals have higher reasonable costs than smaller hospitals; and Section IX summarizes evidence as to whether the current case mix adjustment system is sufficiently precise to support national payment rates without a hospital-specific component.

I. THE PPS MODEL OF INTERHOSPITAL CASE MIX DIFFERENCES

In Chapter 1 we described the way in which PPS pays hospitals. PPS procedures can be restated as a theory or model of how costs should differ across hospitals. PPS implies a model in which five variables (labor costs, teaching, urban/rural location, case mix, and outlier cases) should determine hospital costs and, therefore, will determine Medicare payments. In addition, the model implied that actual costs and geographic region were to be considered only to cushion the shock of transition and were not allowable sources of difference in the long run. In theory, these variables determine payment through a purely mathematical and objective analysis of recent hospital costs. In fact, a number of judgments were necessary to shape the payment formula. Examining these variables and the judgments which were made clarifies how an implicit model of hospital differences results in a particular payment policy:

A. URBAN/RURAL LOCATION. The urban/rural adjustment is based on the empirical observation that per-case costs, after adjustment for wages, CMI, teaching intensity, and bedsize, are substantially lower in rural areas. Although this difference might reflect lower costs of doing business which have not yet been identified, more likely causes are use of simpler technology and treatment methods and referral of more difficult and costly cases to urban medical centers. The designers of PPS made several decisions in addressing this problem:

1. They decided to treat location as a dichotomous variable rather than as a

continuous variable (based, for example, on population of the MSA in which the hospital was located). This decision was made despite the fact that the econometric equations on which the calculations were based used a four-level measure of location. The most important reason for this decision was the complexity of maintaining multiple levels and the evidence that no simple continuous variable would work well. But even if the underlying typology of urban and rural hospitals is appropriate for payment, there is very likely to be a continuum between the two types. Since the standard payment amount is higher in urban areas, some hospitals classified as urban may be paid too much because they are on the border of being rural and some which are classified as rural will be paid too little because they are on the border of being urban (Table 4.1A). This boundary problem is one of the more difficult technical issues in PPS.

2. A number of small but relatively sophisticated rural hospitals pointed out that they would do very poorly under this dichotomy, so Congress established the concept of regional referral centers, which would be paid at urban rates despite their location (about 160 hospitals are so designated).

3. On the other extreme, a number of small rural hospitals argued that they did not have the flexibility to adapt to national rates, so Congress established the concept of sole community hospitals, which would retain 75 percent hospital-specific payment and would not be moved to national rates (about 360 hospitals are designated as sole community hospitals.

In addition to further discussion of this issue in this report, the Department will deliver to Congress a separate, comprehensive report on the urban-rural issue.

B. LABOR COSTS. Labor costs are measured by a direct survey of wages. We focus here on the the calculation made to determine the relation of the area wage level to hospital costs, and hence to Medicare's payment rate.

Table 4.1A shows that, controlling for case mix index, hospital size, location, and teaching, hospital costs rise 0.84-1.02 percent for every 1.0 percent increase in the area wage level. The architects of PPS calculated that local wages influence 79.15 percent of hospital costs (about 60 percent in direct labor costs and about 20 percent more in purchase of local services which are influenced by local labor costs) and decided that 79.15 percent of PPS payments would be adjusted for local labor costs. This meant a 0.79 percent rise for every 1.0 percent increase in the area wage level.

The reason that cost per case varies more with the area wage level than the calculated 79.15 percent is that the area wage level is also a proxy for other factors associated with higher hospital costs which are not measured directly. Those factors probably include additional operating expenses related to urban location and an association of high labor costs with a more intense style of medical practice. A clear example of the way in which the wage index can serve as a proxy for other factors can be seen by comparing Table 4.1F with Table 4.1B. Adding urban/rural location to the equation decreases the apparent relation between cost per case and wage index. When urban/rural location is not considered, the cost of a case increases 1.43 percent for every 1.0 percent increase in the wage index because the lower wages in rural areas are also serving as a proxy for the less expensive practice patterns in rural areas. When the relation of cost to wage index is adjusted for differences between urban and rural location, the cost of a case increases only 0.98 percent for every 1.0 percent increase in the wage index.

The designers of PPS chose, based on evidence as to what the effect of wages on cost per case should be, to regard the remaining effect as a kind of variation for which Medicare should not pay. They did not choose to use the wage index as a proxy for unmeasured case mix variation. The decision to apply the local wage index to 79.15 percent of hospital costs was intended to reflect the actual way hospitals purchase labor and labor-related services, and to avoid a perverse incentive to hospitals to raise their wages. A secondary effect, however, was to make PPS a somewhat less exact model of actual differences between hospitals.

C. TEACHING ACTIVITY. Teaching activity is measured by the ratio of the total number of interns and residents to the total number of beds in a hospital. But the way that this measurement is used to determine payment again reflects judgment. When the architects of PPS calculated the effect of teaching on costs they simultaneously calculated the effects of hospital size (measured by total beds). When they controlled for hospital size, the effect of teaching on cost per case was much lower than when hospital size was omitted from the calculations. That is, some of the higher cost seen in large teaching hospitals was also seen in large non teaching hospitals.

The architects of PPS decided for a variety of reasons that the tendency of increasing hospital size to increase cost per admission was probably due to inefficiency; there was little empirical evidence either to support or to refute their decision. Because they did not wish to encourage large size, they controlled for hospital size in calculating the indirect cost of teaching but excluded size from the payment system. They therefore adopted the measure of teaching effect which had been calculated after controlling for the effect of size--about a 5.8 percent increase in cost per case for every tenth of an intern/resident per bed (Table 4.1A). Had they excluded bedsize from the calculation of the teaching adjustment, they would have found an 8.5 percent increase in cost for every tenth of an intern/resident per bed (Table 4.1B). Although the magnitude of these adjustments is somewhat different in calculations for 1984, the relationship is similar. Teaching hospitals argued to

Congress that they would lose money under the 5.8 percent adjustment; Congress responded by doubling the initially proposed teaching adjustment to 11.6 percent. Thus Congress overcompensated for the decrease which had resulted from controlling for the effect of hospital size.

The designers of PPS did not attempt to determine whether the increased costs in teaching hospitals related to the cost of teaching, the cost of caring for sicker patients, the cost of more intensive treatment, or inefficiency. Through a contract from the Office of the Assistant Secretary for Planning and Evaluation, the Department has just completed the Graduate Medical Education study, which examines the sources of these cost differences. Analyses are just being completed and are being published separately.

There is also evidence that the indirect teaching adjustment has a more complex relation to the number of interns and residents per bed than the simple percentage increase. The indirect teaching adjustment actually is somewhat too high for hospitals with high teaching activity compared to those with low activity. For example, Cretin et al. (1986) show that, at national rates, profitability of teaching hospitals rises twice as rapidly with increased teaching for hospitals with 0.25-0.49 residents per bed as for hospitals with 0.01-0.24 residents per bed (see also Sheingold 1985 and Cotterill et al., 1985 for comparable results from other data). In COBRA, Congress legislated changes in the teaching adjustment which bring payments slightly closer to this nonlinear relationship.

D. DRG MEASUREMENT OF CASE MIX. In Section IV of this chapter we discuss the evidence that DRG weights are "compressed"--that is, weights for complex cases are too low and weights for simple cases are too high. If weights are compressed, a hospital's average cost per case will rise faster than its CMI.

Whether weights appear compressed depends, however, on the way in which hospital costs are predicted from independent variables. Tables 4.1A and 4.1B show that, again, whether weights appear compressed depends on whether the calculation includes hospital size. If the equation includes residents/bed, wage index, CMI, and urban/rural differences, then cost per discharge goes up about 12.8-14.6 percent for every 10 percent increase in the CMI, indicating compression. But when hospital size is included, the increase in cost is almost exactly 10 percent for every 10 percent increase in the CMI. Thus, including the effect of hospital size when estimating the payment corrections and excluding it from payment because of a suspected relationship to inefficiency also solved the apparent problem of DRG weight compression. We will return to the hospital size issue in Section VII.

E. OUTLIER POLICY. At first examination, outlier policy was based on a wholly rational economic decision that the outlier system would pay only the marginal costs of additional care and by using the best available estimates of such costs. More careful examination makes clear that the apparently mathematical and economic decisions were shaped by policy judgments as much as by numbers, as two examples will illustrate:

1. Marginal Costs for Cost Outliers. In implementing the PPS law, marginal cost is defined as the additional costs of care beyond the outlier threshold. There is no evidence that the additional cost of services provided to cost outliers beyond the threshold is different from the cost of services provided to these cases before the threshold is reached. Data available when implementing regulations were written suggested that marginal costs for day outliers were probably lower because there were substantial fixed costs associated with extra days of stay, and this evidence was extrapolated to cost outliers. Section VI discusses current evidence on the marginal cost of day and cost outliers.

2. Limiting Outlier Payments to the Federal Share. The architects of PPS decided to apply outlier payments only to the Federal portion of the payment (25 percent in the first year) and not to the hospital-specific portion (75 percent in the first year). To the extent that outlier policy was intended to protect expensive hospitals with a disproportionate share of outliers, this was logical because the hospital-specific portion of the payment rates incorporated the costs which the hospital incurred from having a large number of outliers in the base year. But outlier policy was also intended to provide insurance against the risk of occasional very expensive cases for smaller hospitals with average costs. The small hospital needs this insurance because it does not have enough discharge to "average out" exceptional cases. The fixed hospital-specific portion of the payment causes as much risk as the fixed Federal portion, but the insurance applied only to the latter.

These two decisions on outlier payment were shaped by a common policy goal, which was to assure that while the outlier payment system reduced the risk associated with outliers it would under no circumstances make outliers fiscally attractive.

F. DISPROPORTIONATE SHARE OF MEDICARE, MEDICAID, AND LOW INCOME PATIENTS. The original PPS legislation authorized the Secretary to find a way to make special adjustments in payments to hospitals with a disproportionate share of Medicare, Medicaid, and other low income patients. Recent legislation (COBRA) establishes specific formulas for increasing payments which are based on the proportion of a hospital's Medicare patients who are eligible for Supplemental Security Income payments and the proportion of non-Medicare patients who are eligible for Medicaid. This adjustment does not necessarily rest on needs of individual patients, but to the extent that PPS is built on a case mix model, Congress thus implicitly accepted poverty as a legitimate proxy for case mix variation which was unmeasured by the CMI and the outlier adjustment. HCFA is examining whether this difference is actually observable under PPS.

G. REGIONAL ADJUSTMENTS. Regional differences in hospital treatment patterns are well documented and remain a perplexing problem for Medicare policy. The assumption in the law is that, in the long run, they need not be recognized--perhaps because they are thought to result from unnecessary regional variations in efficiency or treatment practices; the regional component of payment rates is scheduled to phase out by 1989 (see Table 1.1). These differences are not at all well understood and might result from differences in patterns of illness or hospitalization which should be considered in setting payment policy; there are, however, no data to support such explanation.

H. CONCLUSIONS. This analysis shows how apparently mathematical decisions required policy choices, which often produced payment rules which were intended to be efficient rather than those which would most closely reflect current medical practice. Thus, significant reallocations of Medicare payments among types of hospitals, among regions of the country, and between typical and atypical patients were built into PPS by its designers. Although the circumstances which shaped those decisions may have changed somewhat in 3 years, and though we now have more data with which to assess whether they should be continued, it is important to keep in mind that the goal of PPS was neither to perfectly duplicate existing payment patterns nor to pay costs. The architects of PPS, both in the Administration and in Congress, clearly intended to go beyond those standards and to pay what they judged to be the costs of delivering necessary care in hospitals of average efficiency.

II. HOW PATIENTS WITH HIGH TREATMENT NEEDS CONCENTRATE IN SOME HOSPITALS

At this point, we change focus slightly and consider how interhospital differences in case mix actually occur. Since urban/rural location and teaching activity are only proxies for patient needs, we need a model of how patients with greater needs might concentrate in some hospitals. If all hospitals received a random selection of patients, we would expect only random or nonsystematic variation in case mix among hospitals. Different hospitals might thus show very similar patterns of variation within the uncomplicated myocardial infarction DRG if all patients with myocardial infarction were simply taken to the physically nearest hospital.

However, different hospitals appear to have patients with different average care needs, and these differences appear to result from differences between hospitals along three dimensions:

--who lives in the hospital's catchment area,

--how patients are referred to different hospitals, and

--how hospitals define the needs of patients.

A. CATCHMENT AREA FACTORS. An obvious source of difference between hospitals is differences between catchment areas which they serve. Thus, a hospital serving a catchment area with an excess of the very poor, nursing home residents, or alcoholics is likely to have a patient population with different needs from a hospital with an average catchment area population. Measuring these differences is difficult, and we are not aware of research, either complete or in progress, which would permit relating catchment area factors to hospital level of severity.

B. REFERRAL PATTERNS. Hospitals acquire a particular mix of patients, which differs from that of other hospitals with similar catchment areas, through a variety of referral systems.

1. Reasons for Referrals. The reasons for referrals to particular hospitals include hospital characteristics such as:

--availability of diagnostic or therapeutic services,

--perceived superior or inferior services,

--perceived willingness or unwillingness to accept patients in certain medical or financial conditions, and

--desire to take advantage of or avoid a style of practice which is characteristic of a hospital.

2. Mechanisms for Referrals. The mechanisms of referral also vary widely and include:

--hospital-to-hospital transfers;

--emergency room transfers;

--selection of hospitals by physicians, especially physicians with privileges at more than one hospital who admit selectively or differentially;

--referral agreements with institutions such as nursing homes and community health centers;

--ambulance runs, emergency medical service networks, and similar transport systems; and

--self-referral by patients.

Such referral mechanisms produce concentrations of different kinds of patients in different hospitals. To the extent that such concentrations are based on clinical judgment by physicians, classification systems are unlikely to capture the impact on hospital case mix because clinical referrals are likely to be made for more subtle reasons than available patient data can capture.

C. HOSPITAL DIFFERENCES AND DIFFERING DEFINITIONS OF PATIENT NEEDS.

The increased cost per case observed in teaching hospitals is often attributed to patients being sicker or needing more intense care. In Chapter 2, however, we pointed out the limitations of the "severity" model of patient needs. We begin here to develop a more complex model of hospital differences which builds on our knowledge of the limits of the severity model and which appears to take better account of known facts. To illustrate this model, we will describe a hypothesis as to how patients differ between major teaching and community hospitals.

Major teaching hospitals can be described, for this hypothesis, as having two kinds of care.

1. Specialty Units. This care is concentrated in highly specialized, acute referral units, such a trauma, burn, neonatal intensive care, and immune deficiency services. In these units, which represent a small fraction of teaching hospital admissions, severity of illness is often very high and costs are very high as a result of this severity. Costs maybe much higher for patients admitted to such units than for patients in the same DRG not admitted to such units.

2. Routine Hospital Operations. This care includes patients referred for diagnostic and special treatment services, but generally on a nonemergency basis. These patients are not more severely ill than comparable patients in community hospitals, but a large fraction of them receive more intense care for a variety of reasons:

--They are referred for subtle or highly specialized diagnostic evaluation or for highly specialized judgment as to appropriate treatment.

--They, their doctors, and/or their family want everything possible done to preserve life or functioning. Here the existence of a house staff which will do everything possible at 3 in the morning is a reason for referral. By contrast, many physicians might prefer to admit a clearly dying patient to a community hospital if only supportive care and relief of suffering were sought.

--A significant number of referred patients arrive with early disease but hope for curative treatment. These patients are precisely those who confound the severity model by receiving care which is out of proportion to the severity of their illness.

--The hospital is organized to identify the need for and to carry out a larger number of therapeutic interventions for routine and unreferred patients than is a community hospital. Although the appropriateness of these interventions tends to be supportable in terms of the medical literature, there is no good literature to suggest that matched patients have better long-term survival from university hospitals than from community hospitals (see Garber et al. 1984 for confirmatory data).

The exact explanation for the differences between high intensity and low intensity hospitals is very complex and involves at least two separate issues: technology diffusion and induced demand.

3. Technology Diffusion. New procedures and new methods necessarily are introduced in a few facilities and, if successful, diffuse outward to other facilities. Most of the time, these innovations begin at large hospitals, usually teaching hospitals, and move outward from these centers both geographically and intellectually.

4. Induced Demand. For some of new procedures, patients will come to the centers of innovation, but in many other cases the procedures will simply be available to those patients who come to the centers for other reasons. For these latter patients, economists might say that the availability of the procedure induces demand. But it is important to distinguish the kind of induced demand in which the service is clearly medically appropriate from induced demand where the need for the service is uncertain. There is considerable debate about whether large, teaching

hospitals provide large numbers of unnecessary services and, if so, whether this occurs because of induced demand or because of inexperienced house staff who order unnecessary services.

The idea that different hospitals may legitimately define patient needs differently creates profound problems for the severity model of patient needs. It also raises some question as to whether increased costs in larger hospitals can be entirely dismissed as the result of inefficiency, or whether larger hospitals, with more consultants and more equipment, may share some characteristics with teaching hospitals.

III. EMPIRICAL EVIDENCE ON THE DIFFERENTIATION OF HOSPITALS

One critical empirical question regarding the severity model is whether it can account for differences currently dealt with by the urban/rural and teaching adjustments. A severity measure need not justify current payments for teaching or current urban/rural differences, but it should substantially clarify the contribution of severity to the increased costs of urban and teaching hospitals.

The difficulty in analyzing differences effectively is that we have no data set which directly measures clinical severity of illness in different types of hospitals even if we accept such measures as APACHEII, MEDISGRPS, and SOII as valid measures of severity. We have no measure of medically appropriate intensity of care independent of severity. Thus, all evidence regarding patient differences across classes of hospitals is somewhat indirect.

A. DIFFERENCES IN PATIENT TREATMENT BETWEEN TEACHING AND NONTEACHING HOSPITALS. Good research on differences in treatment between teaching and nonteaching hospitals is sparse. Garber et al. (1984) compared treatment on teaching and nonteaching services in a university hospital. They found that patients in the teaching service had higher costs, higher survival rates in hospital, but equal survival rates at 9-month followup compared to carefully matched patients on the nonteaching service. These data strongly suggest, but do not prove, that more intensive therapeutic efforts were made on the teaching service and that the more intense effort resulted in higher cost and better short-term survival but no change in long-term survival.

Cretin et al. (1986) report that the correlation of profitability of individual DRGs in nonteaching and teaching hospitals is 0.61-0.74 (R^2=0.37-0.55). This means that about half of the variation in profitability among DRGs in teaching hospitals is reflected in the profitability of the same DRGs in nonteaching hospitals. This suggests that teaching hospitals are not simply doing more than nonteaching hospitals; they are doing different things in different patterns which are diagnosis/related.

B. DIFFERENCES IN CASE MIX BETWEEN TEACHING AND NONTEACHING HOSPITALS. Examination of the CMI and outlier payments reveals that case mix as measured by PPS is significantly higher in teaching than in nonteaching hospitals. But a further question is whether case mix in teaching hospitals is actually even higher but not fully measured by PPS. Such an unmeasured case mix difference would provide strong justification for the higher costs of teaching hospitals, which are reflected in the teaching adjustment. Studies by Coffey and Goldfarb (1984), Coffey (1985), Short and Coffey (1984), and Mitchell et al. (1985b) tell a story which is relatively consistent and coherent.

Coffey and associates of the National Center for Health Services Research and Health Care Technology Assessment, have conducted a number of studies which used Disease Staging to examine the differences between teaching and nonteaching hospitals. Disease Staging (Chapter 3.III) classifies patients according to the progression of the disease--stage 1 is an early highly localized condition and stage 4 is death. Although Disease Staging depends on discharge abstract data and does not include any physiologic measures of severity of illness, it should indicate whether patients have more advanced disease and, through use of death, should provide a measure of severity. Coffey and colleagues found consistently that Disease Staging did not demonstrate a significant difference in case mix between teaching and nonteaching hospitals.

Coffey and Goldfarb and Mitchell et al. (1985b) did find consistently greater use of procedures, including surgery, in teaching hospitals compared to nonteaching hospitals. Their data did not permit them to distinguish patients referred to the teaching hospital for specialized surgery from those who were simply routine users of the teaching hospital but were more likely to have surgery for their conditions. Thus, the greater use of procedures in teaching hospitals appears not to be explained by the relatively crude adjustment provided by staging.

C. TREATMENT PATTERNS ACROSS REGION. Eggers (unpublished) has provided a striking analysis of the different ways in which hospitals are used in different States. He compared the number of patients dying in hospitals with the number of patients dying within 30 days of a hospital admission (Table 4.3). (This number can be negative if more patients die after a hospital stay of more than 30 days than die outside of the hospital during the 30 days following admission.) The differences across States are dramatic and suggest how differently hospitals fit into the overall health care system in different States. These data provide, however, no guidance on the extent to which differences reflect differences in quality of care as opposed to differences in the way the system cares for dying patients.

Mitchell et al. (1985b) used a somewhat different method. Using Medicare data from four States, they examined clusters of DRGs associated with cerebrovascular disease, coronary artery disease, prostatic disease, and pneumonia. They also used Disease Staging as a measure of disease severity within DRG cluster and used the presence of specific diagnoses to further refine their estimates. Again, these measures accounted for very little of the difference either in cost or in procedures performed. It is difficult to imagine any system of referral or any catchment area differentiation which would account for the differences in the rates of performance of certain procedures, such as CAT scans in patients with stroke, which they observed. The data strongly suggest that hospitals in different States provide different services to patients who are more similar than the mix of services would suggest.

Nevertheless other data suggest that differences across States may reflect other factors which would affect both assignment of patients to DRGs and severity of illness within DRGs. First, rates of hospitalization differ sharply across States (Beebe et al. 1985). Second, diagnostic patterns differ across States in ways which might well result in different DRG assignments for rather similar patients (Iezzoni and Moskowitz 1984). Third, frequency of surgery and other procedures differs across States very dramatically (Chassin et al. 1986). Mitchell et al. (1985b) also show that both surgery rates per 1,000 Medicare beneficiaries and surgery rates per admission vary dramatically across the four States they studied (Table 4.4). Fourth, Mitchell et al. find substantial differences in the degree of disease progression across States for a number of DRGs; although Stage is a most imperfect measure of severity, finding substantial differences in Stage across States suggests either differences in patients or differences in diagnostic practices. Finally, Mitchell et al.. show large DRG-specific differences in death rates per admission across States (Table 4.5), suggesting that hospitals are being used for terminal care in some States but not in others. The cumulative effect of these findings is to cast significant doubt on whether hospital stays have the same purposes and whether DRGs have fully comparable patients across States.

D. URBAN/RURAL, TEACHING/NONTEACHING, AND HOSPITAL SIZE DIFFERENCES IN CARE. Cromwell et al. (1985) examined differences in inpatient care controlling for three dimensions of hospital differences: urban/rural, teaching/nonteaching, and large/small. The picture is relatively consistent across the three classes--patients in large, urban, and teaching hospitals tend to get more procedures of all kinds than those in small, rural, and nonteaching hospitals. Their major control for severity of illness was Disease Staging, whose power we have already reported to be fairly modest, but the results are similar to and confirmatory of Goldfarb and associates. Cromwell's findings, like Goldfarb's, are striking not because Staging failed to explain all the difference between hospital classes but because it explained almost none.

The relationship of hospital size to cost of care varies in a way which suggests that the key issue may be differences between medium-sized and small hospitals rather than between medium-sized and large hospitals. Table 4.2 shows that, after controlling for teaching, case mix, urban/rural, and labor costs, cost per discharge was 17.6 percent higher in hospitals of 100-149 beds than in hospitals with less than 50 beds. Average cost per discharge in hospitals over 399 beds was only 6.3 percent higher than in hospitals of 100-149 beds. Cretin et al. (1986 Figures 4.1 and 4.2) found a comparable pattern in which profits varied much more in the lower bed range than in the upper; they also found that profits actually rose in the very largest hospitals, suggesting that costs fell. Thus, the relation of cost to hospital size mainly reflects differences between small hospitals and all other hospitals. These finding are at least consistent with the hypothesis that the small hospitals do less for patients and that patients with greater needs are either transferred or selectively admitted to larger hospitals. Such a pattern is easier to imagine between small and medium-sized hospitals than between medium-sized and large hospitals.

E. DIFFERENCE IN COSTS BETWEEN SETTINGS. Cost differences are strikingly illuminated by Cromwell et al.'s examination of prostate disease, which appears to vary little in either Disease Stage or services rendered across types of hospitals. They found that the pattern of stages for both benign prostatic hypertrophy and prostatic cancer was generally similar across hospital types. The only differences were a somewhat higher staging of hypertrophy in small hospitals and a slightly higher staging of both hypertrophy and cancer in nonteaching and minor teaching hospitals, and in urban hospitals (Cromwell et al. 1985; Tables 7, 13, and 19).

When they examined the mix of procedures for these two conditions they also found a very homogeneous pattern of procedures and a very similar rate of performing procedures in urban and rural hospitals and in teaching and nonteaching hospitals (Cromwell et al. 1985; Tables 8 and 14), but a significant trend to performing more procedures in large hospitals compared to small hospitals (Cromwell et al. 1985; Table 20).

Despite the homogeneity of Disease Stage and procedures, urban hospitals had 18.9 percent higher costs than rural hospitals after adjustment for wage differences (Table 2), major teaching hospitals had costs 70.9 percent higher than nonteaching hospitals (Table 9), but large hospitals, where more procedures were performed, had costs only 6.2 percent higher than small hospitals (Table 15). In these cases, therefore, neither severity of illness nor more intensive use of procedures appears to drive the higher costs; instead, per diem costs are higher in the urban and teaching hospitals, and unit costs for procedures may also be higher.

Clearly, this single study does not explain the differences between classes of hospitals. Work in progress by Iezzoni at Boston University, Ashcraft at University of Michigan, and Alemi at Tulane promises to clarify the picture substantially.

The next section of this chapter examines issues in hospital cost finding which may explain some of the cost differences reported by Cromwell. Briefly, unit costs tend to be "cross-subsidized": hospital cost finding tends to average costs for all patients so that reported costs are, therefore, higher than actual costs for the simplest cases in a hospital and reported costs are lower than actual costs for the most complex cases in a hospital. If this cross-subsidization occurs, it could make moderate cost patients such as those with prostate disease appear to have higher costs than they actually incur in teaching hospitals and appear to have lower costs than they actually incur in rural hospitals.

IV. SETTING DRG WEIGHTS AND THE PROBLEM OF COMPARING HOSPITALS

This section examines two problems in setting the relative weights for DRGs. Part A describes the technical problems in comparing hospitals which result from a phenomenon called "weight compression"; Part B examines the problem of computing suitable weights or prices for making comparisons between hospitals.

A. DRG WEIGHT COMPRESSION. Pettengill and Vertrees (1982), and more recently Lave (1985) and Pauly and Wilson (1985), have examined evidence that DRG weights might be "compressed"--that is, that the payment weights may be too high for simple DRGs and too low for complex DRGs. Compression in the weights would be likely to result in payments which were too high for hospitals with a simple case mix and too low for hospitals with a complex case mix. Four forces can produce compressed weights: differences in cost patterns, data errors, cross-subsidization in per diem cost finding, and cross-subsidization in ancillary charges.

1. Differences in Cost Patterns. Payments are based on relative costliness of

patients in different DRGs in all PPS hospitals. This means they are averages of costs, and hence of practice in different kinds of hospitals. Averaging two or more sets of DRG weights which are not perfectly consistent will result in average weights which are more compressed than any of the sets of weights which were averaged, a statistical phenomenon termed "regression to the mean."

DRG weights derived from different kinds of hospitals and different States are not perfectly correlated. Although Pettengill and Vertrees (1982) found similar weights across regions, they found substantial variation in patterns of weights across hospital types. Lave reports significant differences between the cost weights for New Jersey, Maryland, and Medicare data. Cretin et al. (1986) found that patterns of profitability differ between teaching and nonteaching hospitals, suggesting that cost weights calculated for these two groups of hospitals would differ significantly. All of these findings indicate that national weights based on averages would be compressed. The fact that weights are computed by averaging individual cases rather than sets of weights does not protect against this source of compression.

2. Data Errors. Data errors which result in misclassifying cases or in misreporting costs blur the sharp distinctions between classes of cases, make cases look more like one another than they really are, and result in cost weights which are also more alike than they should be--i.e., compressed. Data errors appear to have been significant in the data used to establish weights (Worthman and Cretin 1986). Although Lave (1985) argues that PPS will correct this effect through increasing attention to coding, and Carter and Ginsburg (1986) have provided evidence that such correction is taking place, the role of data error in current weights cannot be established.

3. Cross-Subsidization of Per Diem Charges. Lave (1985) points out that the basic Medicare cost finding mechanism produces uniform routine per diem and

special care per-diem costs for all cases and that this inherently produces cross-subsidization between cases with intensive per diem care needs and those with less intensive needs. A study by the American Nurses' Association (McKibbin et al. 1985) indicates that such cross-subsidies do occur and current work by Thompson et al. (1984) supports these findings.

4. Cross-Subsidization of Ancillary Charges. In contrast to per diem cross-subsidization, which is partly the result of the Medicare cost-finding procedure, cross-subsidization of ancillary charges results from inaccurate charging by hospitals. Evidence for such cross-subsidies is much weaker than evidence regarding per diems, and Lave argues cogently that such cross-subsidies are likely to disappear under the pressure of competition, which will generally force providers to price individual services more accurately in order to remain competitive.

This analysis makes clear that DRG weight compression does occur. The hypothesis can be tested in fairly simple fashion by determining whether hospital cost per case changes more than 10 percent for every 10 percent change in the CMI. Table 4.1 shows that, in 1981, average cost per discharge changed 28.91 percent for every 10 percent change in the CMI; after controlling for teaching intensity, urban/rural, and the wage index, average cost changed 14.6 percent for every 10 percent change in the CMI; and controlling for hospital size reduces the change to 10.1 percent for every 10 percent change in the CMI. Thorpe et al. (1986) report that in 1984 estimated cost per case rose 12.8 percent for every 10 percent rise in the CMI, after controlling for teaching, urban/rural location, and wage index, and region, but that the increase was only 10.6 percent for every 10 percent rise in the CMI when they included hospital size in the predictive equation. Thus, evidence of compression is present unless hospital size is included in the equation predicts cost per case. We have already pointed out that the practical importance of compression is that hospitals with complex case mix would be underpaid and hospitals with simple case mix would be overpaid. To examine this issue, Thorpe et al. "decompressed" DRG weights in such a way that the average weight did not change but weights above the average were increased and weights below the average were decreased. This approach also decompressed the hospital CMIs. The effect is to reduce evidence of compression with negligible impact on simulated program outlays. Thorpe et al. examined the effect of decompression both including and excluding hospital size from the predictive equation. With size excluded, they found that decompression would reduce the indirect teaching adjustment by 12.2 percent; with size included, compression was very small and decompression had a negligible effect (Table 4.6). These findings suggest that the decompression strategy used in Thorpe's model is an oversimplification of the actual nature of compression. Apparently, some weights are more compressed than others and the compression of weights is largely concentrated in DRGs which have higher costs in teaching than in nonteaching hospitals and higher costs in large hospitals than in small hospitals. We would expect this effect if compression actually results from averaging case mix patterns which differ across hospital types.

Thus, paying on hospital size and teaching intensity would eliminate the effects of compression because these two variables can accurately reflect and correct for the effect of compression on hospitals. These findings suggest that decompressing the DRG weights would result in increased payments to large hospitals and teaching hospitals. We will return in Section VII, to whether decompressing DRG weights would be a useful way to respect the effect of hospital size on costs without actually basing payments on hospital size.

B. SELECTING AN APPROPRIATE DATA BASE FOR CREATING WEIGHTS. DRG weight compression might result in payment errors across hospital types. Other kinds of weighting and data errors could simply reduce the apparent accuracy of the payment system without introducing systematic biases against a particular group of hospitals. In Chapters 2 and 3 we pointed out that using actual costs as a proxy for necessary care risks errors; Section 1 of this chapter pointed to two areas (hospital size and outliers) in which the designers of PPS determined that actual costs would not be used as a proxy for appropriate costs. This raises the question of whether unnecessary or misestimated costs included in the data base have an overall distorting effect on DRG weights, as the designers of PPS thought they might.

1. Inclusion of Inappropriate Care in the Data Base. Gertman and Restuccia (Gertman and Bucher 1971, Gertman and Restuccia 1981, Restuccia et al. 1984) and Systemetrics (1983) have shown that a substantial amount of care is delivered in acute hospital beds which does not require hospitalization. Such inappropriate use means that current patterns of care do not represent necessary care. Some evidence also suggests that medically appropriate services are not always rendered, and that costs are, therefore, understated for some conditions. On the other hand, Kay et al. (1986) found evidence (mentioned in Chapter 3) that these errors may not distort either weights nor estimates of the explanatory power of DRGs. Kay et al. found that removing days of care which did not require hospitalization (nonacute days) from the data base changed neither the coefficient of variation for individual DRGs nor the average length of stay for these DRGs. Since costs were not available and since the reliability and validity of the instrument used have not been entirely proven, this evidence is quite preliminary. Nevertheless, the findings suggest that variations in the appropriateness of care rendered may not be a significant problem in either creating weights or in assessing the power of the system.

2. Problems in Pricing Care. Evidence cited in our discussion of weight compression suggested that determining accurate prices for DRGs could be very

difficult. Langenbrunner (1986) gives an example suggesting that this problem might become serious under some circumstances. Current evidence suggests that the use of flat per-diem rates in general hospitals results in systematic subsidization of pediatric services by other types of services. This appears to occur both because pediatric units have characteristically low occupancy rates with high rates of fluctuation and because pediatric nursing requirements are more intensive than nonpediatric requirements. The result is that Medicare weights for pediatric DRGs are probably systematically understated, with a very slight complementary inflation of other weights. This is not a problem in paying general hospitals, where the errors appear to balance out, but might result in serious underpayment to pediatric specialty hospitals, where there are no nonpediatric beds to subsidize the pediatric services. A similar phenomenon may occur in pricing psychiatric services, where the per diem costs of acutely ill patients are averaged with those of the chronically ill or custodial patients whose care is not part of the Medicare benefit package. In contrast to issues of weight compression and "dirty" calibration data, which appear to have limited significance, distortion of weights by cross-subsidization may have a significant impact on the equity of payment under PPS.

V. THE "ACCURACY" OF PAYMENTS IN THE FIRST YEAR OF PPS

An important source of evidence on hospital-level factors affecting case mix is the actual relationship of PPS payments to costs as well as the resulting profits and losses. Gains or losses by classes of hospitals suggest either that these classes of hospitals have systematic management strengths or weaknesses or that factors for which PPS does not fully account have an important impact on the system's equity. Thus, gains and losses across hospitals that are not consistent with the objectives of PPS would invite a reexamination of whether PPS accurately reflects the real world in which hospitals operate.

The accuracy of PPS is usually measured by how closely payments match current costs, but this criterion must be very carefully applied. The possibility of greater efficiency is one of the foundations of PPS, so a perfect match between costs and payments is not expected and would probably not be desirable. On the other hand, much of PPS is an effort to model current costs fairly accurately, so substantial correspondence between costs and payments is expected and desirable. However, the analysis already presented makes clear that we cannot determine technically the exact degree of correspondence which PPS should attain.

This section analyzes the relation between costs and payments in four ways: the amount of variation in costs which is reflected in payments, the operating margins which various groups of hospitals experienced, the number of hospitals in different groups experiencing various operating margin levels, and the relative association of different hospital characteristics with level of operating margin.

A. METHODS. Because this chapter focuses on hospital-level issues, all comparisons will be made for hospital average costs and payments, not for costs and payments at the individual patient level. Operating margin at the hospital level means the difference between costs for Medicare patients and payments made by Medicare, expressed as a percent of Medicare costs. Operating costs and margins exclude both capital payments and payments for the direct expenses of medical education, which are "passed through" outside the PPS per-case payment mechanism.

All analyses in this section are based on actual Medicare cost reports for the first year in which hospitals were paid under PPS. These cost reports were used in making projections of the impact of the PPS payment regulations for Federal fiscal year 1987. The file contains usable cost and payment information for 4544 hospitals and cost information alone for 5350 hospitals; only a small fraction of these reports had been finally audited and settled when these calculations were made.

In order to project actual payments under a fully implemented PPS system, we took the national component of 1987 payments, which will actually comprise 75 percent of payments in hospital reporting periods beginning in 1987, and assumed a full implementation in which this national component was 100 percent of payments (except for sole community hospitals). We further adjusted to reflect the shift in calculating urban and rural rates which will become effective in 1988. We then applied that payment rate to discharges during the hospital's first PPS year. For estimates of costs at full implementation we inflated costs from 1984 at the inflation rate of HCFA's hospital market basket. The reader should notice that though we have projected the situation under full implementation we have no knowledge of what changes may occur in regulations between 1986 and 1988, of the actual rate of inflation, or of changes in hospital productivity between 1984 and 1988. We have applied the 1987 "disproportionate share" adjustment and the adjustments required by the Omnibus Budget Reconciliation Act of 1986 (OBRA-86 PL 99-509).

We used estimates developed by Guterman (1986) of the degree to which different classes of hospitals were able to reduce their operating costs in response to PPS. His estimates reflect the difference between the cost per case which HCFA projected for each hospital in establishing hospital-specific payment rates for the first year of PPS and the actual costs reported by hospitals for that year. Because HCFA projected 1984 costs from unaudited 1982 cost reports and overestimated inflation, the resulting cost projections are somewhat high (PROPAC 1985). This means that our estimates of operating cost reductions are also somewhat high; we discuss these issues further in Section E.1.b.

B. OVERALL LEVELS OF MARGINS. We present overall margin levels in several ways so that readers will appreciate the range of estimates which can be made from a single set of data and so that they will not give excessive weight to differences of a few percentage points in margin levels reported by different sources. We can calculate:

- --the operating margin for the entire hospital industry. To do this we compare all PPS operating payments with all Medicare operating costs; by this method, overall margin in 1984 was 16.4 percent and at full implementation would be 13.3 percent
- --the case-weighted average hospital margin. To do this we compute a margin for each hospital and then calculate an average in which each hospital's margin is weighted by the hospital's number of Medicare discharges. This method gives 15.8 percent in 1984 and 16.0 percent with full implementation.
- --a simple average of hospital margins. To do this we compute a margin for each hospital and calculate an average in which each hospital has equal weight regardless of size. This gives 11.6 percent in 1984 and 17.7 percent with full implementation.

Each of these figures has advantages and disadvantages. The industry figure reflects the relation of what Medicare is paying to the cost of producing the care it pays for; it is least vulnerable to auditing and data errors. The case-weighted average gives greater weight to the experience of a large hospital than to that of a small hospital, while the hospital-weighted average gives equal weight to each hospital; both case-weighted and hospital-weighted averages are quite sensitive to data errors and to how extreme data values are trimmed. For example, both averages show higher margins at full implementation than in 1984 while the hospital-level figure shows the reverse pattern. This occurs because very high profits have been trimmed from the 1984 data but not from the full implementation projections. By any computation, however, margins are higher than many had expected at the beginning of PPS.

1. Reasons for High Margins. Several forces appear to have created the high overall margins:

a. Overestimated Base Year Costs. In order to use the most recent data, HCFA employed unaudited cost reports to set payment rates; such reports characteristically overstate costs or include costs which HCFA ultimately does not allow. A General Accounting Office (GAO 1985) analysis estimated that payment rates were 4.3 percent higher than they would have been if based on audited reports.

b. Coding Changes which Increased Payments. Upgraded coding resulted in higher average case mix index (CMI) than had been anticipated. Carter and Ginsburg (1986) conclude that the higher CMI was largely attributable to secular trends and the more accurate case mix measurement made possible by the new Medicare billing form. They also found, however, a modest but sharp increase in case mix at the time hospitals went on PPS. They provide evidence that this increase was a one-time phenomenon; other studies of coding error suggest that the observed change was smaller than would have been expected if previous coding errors had been entirely corrected. Carter and Ginsburg's analysis suggests that this change in coding resulted in 4.9 percent higher margins than would have otherwise occurred.

c. Overestimated Inflation. In setting the base rates, HCFA estimated a higher inflation rate for the hospital market basket than actually occurred.

d. Increased Efficiency. Finally, hospitals appear to have become more efficient. They have decreased length of stay for Medicare patients, decreased staffing, and taken other steps to provide care at lower costs. They have also transferred Medicare costs to skilled nursing facilities and ambulatory settings, producing major decreases in hospital costs but less certain net savings to Medicare.

These changes are documented in the Annual Impact Reports for 1984 and 1985.

2. HCFA and American Hospital Association Estimates. Our estimates for Medicare margins differ from those of the American Hospital Association (AHA), which estimated total net margins of 6.2 percent in 1984 and 5.9 percent in 1985 and net patient margins of 2.0 percent in 1984 and 1.5 percent in 1985 (Hospital Research and Educational Trust 1985). This difference between HCFA and AHA estimates reflects important differences in definition and in the services described:

a. Definitions of Margin.

-- AHA's "net patient margin" includes costs from activities other than patient care but excludes revenues from these sources. AHA's "total net margin" therefore corresponds better to usual notions of profit and loss and is a better basis for comparison with HCFA figures.

-- Medicare calculations of margins are limited to inpatient services to Medicare beneficiaries. By contrast, AHA figures include all hospital activities (inpatient, outpatient, and nonpatient care) and all groups of patients. Although ancillary nonpatient care may be profitable, outpatient business is traditionally unprofitable.

-- Medicare calculations exclude capital and direct medical education from both costs and payments while AHA includes them. Because Medicare pays both of these items at cost, including them in both costs and revenues necessarily brings the calculated margin nearer to zero.

-- AHA calculates margins as a percentage of revenues while HCFA calculates margins as a percentage of costs. When revenues exceed costs, as they did by both AHA and HCFA calculations, the AHA method of calculation always yields a lower margin.

b. Differences in Margins for Medicare and Non-Medicare Patients. For several reasons, profits on Medicare business were higher than profits on non-Medicare business:

- —Medicare paid more than intended in the first year of PPS, which was probably not generally true of other payers
- —Medicare incentives were intended to let hospitals profit by reducing length of stay, eliminating unnecessary services, and providing services more efficiently; other payers sought to control costs primarily by reducing the number of services rendered while continuing to pay on the basis of services actually rendered.
- --Hospital revenue on non-Medicare business is reduced by bad debt and discounts to insurors. Medicare pays for bad debt on Medicare cases and does not demand discounts.

C. HOW WELL DO PPS PAYMENTS REFLECT DIFFERENCES IN OPERATING COSTS? In the first year of PPS, payments reflected 89.0 percent of the variation in costs. When PPS is fully implemented, we estimate that payments will reflect 71.4 percent of variation in 1984 hospital costs (the percent of the variation explained is R^2, which is the square of the simple correlation between costs and revenues at the hospital level). We cannot predict how well payments will reflect the hospital costs which exist at the time of full implementation because we cannot predict what changes hospitals may make in their efficiency between 1984 and 1988 in response to the incentives of PPS or other pressures.

PPS payment variables (urban/rural, teaching, wage index, and case mix index) account for no more of the variation in hospital costs in 1984 than then they did in 1981 (69.2 percent in 1981 v. 68.2 percent in 1984; Table 4.1). The decrease is small but does not suggest that hospitals are becoming more alike through responding to the PPS incentives.

These cost regressions in Tables 4.1 and 4.2 show the limit of how well the PPS payment variables can be used to make payments match costs. The two additional variables not included in the original calculations are disproportionate share payments and outlier payments. Table 4.2 shows that, when they are included, the equation explains 73.7 percent of cost; this roughly matches the 71.4 percent of cost variation which is reflected in payments. The small differences between the R^2 in the regression equation and the actual relation of costs to payment reflects two kinds of differences between the equations and actual payment policy:

- -- the equations do not fully take account of the unique position of sole community hospitals and rural referral centers, and
- -- the payment system is not constructed from the cost equation and, in particular; the equation implies outlier payments about 2.5 times larger than HCFA pays and disproportionate share payments about an eighth of those defined by law. We return to the outlier question in the next section.

The equity of PPS depends on the significance of that cost variation which will not be reflected in payments when the system is fully implemented--an estimated 28.6 percent. If the variation which is not reflected in payments results from factors such as inefficiency, for which Medicare should not pay, then PPS is equitable; if this variation results from factors such as variation in severity of illness within DRGs, for which Medicare should pay, then there is a problem. Unfortunately, we cannot currently determine which of these two factors predominates, so technical procedures cannot fully determine the equity of PPS. Horn et al.(1985a, 1985b) have argued that interhospital differences in severity account for a large part of the difference between hospital costs and PPS payments, but their sample was small and their methods are controversial (Davis 1985, Mullin 1985, Horn 1985c).

Thus a judgment on the equity of PPS payments involves an ultimately nontechnical balancing of the apparent equity of using uniform national payment rates against uncertainty whether uniform payments are truly equitable.

As a result of this technical uncertainty, many critics judge the equity of PPS by how many hospitals are hurt, which ultimately depends on the absolute level of Medicare payments. From this perspective, a more detailed picture of how many hospitals win and lose is illuminating.

D. WINNING AND LOSING HOSPITALS.

1. Magnitudes and Distributions of Margins. Table 4.7 shows average margin by class of hospital. These margins are computed using aggregated payments and costs for the entire group of hospitals. Although overall margins appear strong, Tables 4.8 and 4.9 show that there are significant numbers of hospitals with losses, and even substantial losses, in each group. Overall, 19.5 percent of hospitals had losses in the first year of PPS and 7.4 percent had losses greater than 10 percent of costs. At full implementation, 23.3 percent would have losses and 10.6 percent would lose more than 10 percent. The projected losses under full implementation are, of course, very uncertain because potential losers are very likely to change their practice patterns to avoid such losses.

The range of margins grows under full implementation--that is, there are more big losers and even more big winners: the number losing more than 20 percent goes from 3.0 percent to 3.7 percent (from 160 hospitals to 197) and the number profiting more than 30 percent goes from 9.5 percent to 26.0 percent (from 432 hospitals to 1388). This is expected: PPS was designed with a set of cushions to protect hospitals during transition to national rates, and those cushions will be phased out when payments are fully national. Nevertheless, these findings underline the questions of what the big winners will do with their margins and whether the big losers will reduce access to adequate patient care.

2. Implications of Operating Margins. The impact of a particular operating margin depends both on the relative margins of other hospitals and on the overall changes in margin which may result from future changes in Medicare payment policy. Relative margins are important because hospitals with relatively low margins may have difficulty securing capital and other needed forms of support. Absolute margins are important not only for their immediate dollar value but also as a cushion if payment policy changes. For example, if Medicare payments were adjusted so that overall operating margins dropped from 13.3 percent to 10 percent at full implementation, then hospitals which now have margins less than 3.0 percent would show a loss on Medicare business. Thus, tightening up on the PPS annual update factor would not only reduce average margins but also produce more losers unless hospitals which now have borderline margins could achieve additional economies. Under national payment rates, we project that trimming overall operating margin from 13.3 percent to 10.0 percent would increase the percentage of hospitals with operating losses on Medicare business from 23.3 percent to 32.2 percent of all hospitals.

E. PATTERNS OF MARGINS. Tables 4.7-4.9 show that large, urban, and teaching hospitals did better in the first year of PPS than small, rural and nonteaching hospitals. Table 4.10 shows a regression analysis of the factors which are associated with hospital margins. This table summarizes the effect of each hospital characteristic on operating margin when all other characteristics are held constant. Thus, for example, the difference in margin between urban and rural areas is separated from the effect of different hospital sizes in urban and rural areas. The first point to notice about this analysis is that the associations, although statistically very significant, are relatively weak. The predictors account for no more than 16 percent of total variation in margins either in 1984 or at full implementation. In addition:

1. Rural. An urban hospital would have 10 percent higher profits than a rural hospital under national rates if it had otherwise similar characteristics; but rural hospitals have lower wage rates, which is associated with higher profits. The net effect, as tables 4.8 and 4.9 show, is that urban and rural hospitals will differ little in profitability after adjusting for differences in teaching status and size.

2. Outliers. Including outliers in the variables used to predict profits shows that having outlier cases is associated with losses in 1984 and is much more closely associated with losses at full implementation. Because we cannot tell whether having outlier cases indicates difficult case mix or inefficiency, we rely on the analysis which omits outliers in the following discussion.

3. Efficiency Changes in Small and Rural Hospitals. Table 4.11 shows that both urban and rural hospitals held their costs below the levels which HCFA estimated at the beginning of PPS. As we have mentioned, these "savings" partly reflect HCFA's overestimating costs and partly reflect coding changes, but the difference between HCFA estimates and reported actual costs is far larger than any error which has been suggested and a substantial part probably represents real savings. Rural hospitals achieved an average of 9.5 percent savings in costs during the first PPS year compared to 18.0 percent savings for the average urban hospital. Although this difference in savings may partly reflect the smaller size of rural hospitals, it probably also reflects the fact that rural hospitals were already delivering a leaner service mix (Cromwell et al. 1985). Reducing services may be much more difficult for rural hospitals. As urban hospital costs come closer to those of rural hospitals, deciding whether they should retain the resulting profits will pose tricky problems of equity. A forthcoming report to Congress will address these urban-rural issues in detail.

3. Disproportionate Share Hospitals. Disproportionate share hospitals, contrary to expectations, had slightly higher margins in 1984 than otherwise similar hospitals without a disproportionate share of poor patients. Table 4.0 shows that when these hospitals receive a premium payment, their margins will rise further above average. Disproportionate share hospitals tend to be those (large, urban, and teaching hospitals) which realized greatest efficiencies in 1984. The margins may reflect these savings, since Table 4.2 shows that such hospitals do indeed have slightly higher costs than otherwise similar hospitals.

4. Hospital Size. Table 4.10 shows that hospitals with fewer than 50 beds had slightly lower margins than larger hospitals in 1984 but that these hospitals will have 7-14 percent higher margins under national rates. This increased margin under national rates reflects the fact that the historical costs of small hospitals were about 10-25 percent below those of larger hospitals (Tables 4.1-4.2). In 1984, the difference in historical costs gave small hospitals only a slight advantage because payments were 75 percent based on the historical costs of the individual hospital. The historically lower costs of small hospitals will translate into a substantial windfall when payments are based on national averages under fully national rates. The higher margins of small hospitals reflects lower historical costs rather than increased efficiency since the inception of PPS. Table 4.11 shows that they have achieved less savings than large hospitals under PPS. Section VII examines the problem of hospital size further.

5. Margins in High CMI Hospitals. In 1984 margins were higher at hospitals with high Case Mix Indices (CMI). This relationship disappears in the projection of margins at full implementation. We cannot prove why high CMI hospitals had high margins, but two explanations are plausible:

--Hospitals with a high CMI have more ability to achieve economies. This interpretation is consistent with the economies achieved in large, urban, and teaching hospitals, all of which have higher CMIs.

--Cretin et al. (1985) suggest that CMI "creep" has occurred selectively. If creep was greater in hospitals with CMIs below the CMIs of otherwise similar hospitals but was distributed randomly across hospital types, then hospitals with creep would tend to do better and CMI would be positively related to profit under 1984 rates but not under national rates. Under 1984 rates, the improvement over the historical CMI would produce profit; but a move toward typical CMI for that group of hospitals would not produce greater profit under national rates. This speculative line of reasoning is consistent with the hypothesis that creep consists more in correcting erroneous undercoding than in inflating appropriate coding.

At first glance, these findings do not appear to support the hypothesis that DRG weights are "compressed", since that hypothesis appeared to predict losses in high-CMI hospitals. But we have already shown that the compression of DRG weights manifests itself in overpayments to small hospitals and underpayment to large hospitals, so we expected to find higher margins in small hospitals at full implementation and we would not expect large losses in high-CMI hospitals once we have controlled for differences in hospital size.

6. Teaching Hospitals. Table 4.9 shows and Table 4.10 confirms that teaching hospitals have are substantially higher margins than others and that this trend will persist under projected fully national rates. The greater impact of teaching on margins at national rates is somewhat unexpected because legislation has lowered the teaching adjustment from 11.6 to 8.7 percent. The explanation is that even the lower teaching adjustment is profitable and that teaching hospitals are more profitable when paid at national rates than when paid at their historical costs.

Table 4-5.6 shows that teaching hospitals have achieved very large economies relative to their historical costs; because the teaching adjustment is based on their historical costs, it provides increases in margins.

7. State and Regional Differences. Table 4.12 shows that even after controlling for differences in hospital characteristics there were some differences in operating margins across states and regions in the first year of PPS and that these differences will be substantial under national rates. For example, margins in Mississippi, Kentucky, Tennessee, Alabama, Nebraska, North Dakota, South Dakota, Utah, Idaho. Montana, Colorado, Wyoming, Oregon, and Washington will be 7-14 percentage points higher than margins in the New England States, California, Florida, Illinois, and Michigan. These differences in operating margins primarily reflect differences in costs which existed before PPS and to a far smaller extent reflect differences in efficiencies realized under PPS.

E. SUMMARY. We have already pointed out that we do not have the technical tools which would be necessary to determine whether any of these patterns of profit and loss are equitable. At the same time, some are unexpected and each must be carefully examined to determine whether it conforms to the purposes of PPS. The fact that almost a third of variation in cost is not reflected in variation in payment demands both further research and continued careful monitoring of the impact of the system.

VI. OUTLIER PAYMENTS AND EQUITY TO HOSPITALS

As we indicated at the beginning of this chapter, the outlier payment system was shaped by three forces: a need to protect hospitals from random risk, a need to protect certain hospitals with large numbers of outliers from the fact that the DRG system could not classify such cases, and a need to protect the integrity of the incentives for efficiency which were created by prospective payment by the case.

A. OUTLIERS AND MARGINAL COSTS. PPS legislation defines marginal cost as the basis for outlier payment, but the concept of marginal costs is often puzzling and elusive. The economic definition would be that the marginal cost of the outlier case would be the cost of the case if all the fixed costs of the hospital (such as buildings, equipment, and perhaps some administrative staff) are excluded and the only costs counted are those specific to the patient (such as most staff, dietary, pharmaceuticals, and housekeeping). In some industries, such as electric utilities, the difference between fixed and marginal costs is fairly obvious and fairly easy to define, but this is not true in hospitals. When the original PPS regulations were written, evidence was available to argue that marginal cost represented anywhere from 20 to 90 percent of average costs for all patients. More recent research (Pauley 1985 and unpublished work by Paul Ginsburg of the Rand Corporation) suggests that marginal costs is very close to average cost in the hospital industry. This finding is unexpected in an industry where the cost of equipment and facilities receives so much publicity, and it probably reflects three factors:

1. Actual Hospital Management Practice Has Changed. The 60 percent estimate was based on the best available research, but that was research that was generally 10 to 25 years old. Staffing levels are now much more closely matched to changing workload; empty beds and wards are not left staffed. Because more than 70 percent of hospital cost is labor-related, this change in practice has great impact.

2. Additional Work Often Cannot be Scheduled in Hospitals. Because additional work tends to be the result of medical necessity, it cannot be scheduled and often has the disruptive effect of reducing efficiency instead of using idle capacity.

3. "Fixed" Costs Vary With Volume. Economists usually estimate marginal costs by examining the costs across hospitals as either number of admissions or length of stay increases. But as either of these increases, the sophistication of hospital facilities, cost of capital equipment, complexity of hospital organization, and expense of overhead also tend to increase. Thus, the traditional economic concept of marginal cost becomes difficult to apply.

Thus, research evidence suggests that a new estimate of marginal cost may be needed. Empirical evidence from the first year of PPS bears on this issue as well as the broader issue of the impact of outlier policy.

Guterman (1986) has analyzed the distribution of outlier cases and payments during the first year of PPS. He had information as to hospital charges for each case, Medicare payment (except for direct medical education and capital payments), and hospital and patient characteristics. He estimated operating costs by applying hospital-level cost/charge ratios which had been generated by Cretin et al. (1986) in the Rand synthetic cost study. Omitting capital may make the concentration of outlier payments in large, urban, and teaching hospitals appear slightly larger than it really is. Because the file was complete through April 1985, very few cases are omitted. Guterman identified the difference between charges and payments as a critical variable and examined its distribution across hospitals and beneficiaries.

B. FINDINGS. Guterman found:

1. Outliers account for the majority of the difference between charges and payments under PPS. The 2.96 percent of cases which were classified as outliers accounted for 56.02 percent of the difference between hospital charges and Medicare payments (Table 4.13).

2. Outliers have very high charges and very high losses. Outlier cases are very expensive: 2.96 percent of cases account for 17.66 percent of Medicare covered

charges, 17.19 percent of estimated operating costs, but only 6.35 percent of payments. Average estimated costs for outliers were $16,819, which is 6.80 times the average nonoutlier cost of $2,473. The average payment for an outlier case was $7,248, resulting in an estimated loss of $9,573; this loss contrasts with an estimated profit of $791 on nonoutlier cases (Table 4.14). This estimate is slightly overstated because the payments estimate omits capital and direct medical education payments. Nevertheless, having a large percentage of outliers was, all other things being equal, associated with significantly lower profits in the first year of PPS (Table 4.8).

3. Outliers appear to have more severe illness. Outlier cases appear to be sicker than nonoutliers, even in proportion to their longer stays. Routine daily room charges are 38.0 percent of the bill for the typical patient and 21.3 percent for outliers; intensive care charges are 6.36 percent of typical bills and 11.8 percent of outlier charges; ancillary charges are 55.7 percent of charges for nonoutliers and 66.9 percent for outliers (Table 4.14).

4. PPS day and cost outliers have similar charges but PPS pays more for cost outliers. Despite the very large difference in how day and cost outliers are identified, total charges are rather similar for the two groups: $26,151 for day outliers and $26,012 for cost outliers, as were estimated costs ($17,116 v. $16,385) total. But PPS paid an estimated 51.3 percent of costs for cost outliers and 37.7 percent of costs for stay outliers (Table 4.14).

5. PPS paid less than planned under the outlier payment system. As the 1984 PPS Impact Report noted, the Department's target for outlier payments was 5.7 percent of the "Federal portion" of PPS payments, or about 2.1 percent of total PPS payments (the target amount was slightly more than 5.7 percent of the Federal portion because of delays in publishing a regulation). Actual payments under the

outlier system were about 1.1 percent of total payments, or about 53 percent of the target. Thresholds were changed in 1985 to correct this discrepancy, and the decreasing role of the hospital-specific portion in payments will also increase outlier payments per case.

6. Outliers are concentrated in large, urban, and teaching hospitals. Outlier cases, as well as the corresponding difference between charges and payments, were concentrated in large, urban, and teaching hospitals (Tables 4.14). To some extent, this concentration of high risk cases is offset by the urban/rural and teaching adjustments, but the random risk and the gap between costs and payments are a source of general comment and concern. On the other hand, the combination of cost and day outliers resulted in a relatively even geographic distribution of payments (Table 4.14), with cost outliers concentrated in the western part of the country and day outliers concentrated in the eastern part, which was part of the intent of the design.

C. INTERPRETATION OF FINDINGS.

1. Outliers account for the majority of the difference between charges and payments under PPS. Guterman's analysis shows that this difference comes from four sources:

- Medicare always pays less than charges. Hospital charges always exceed Medicare payment because PPS payments reflect only allowable operating costs, and hospitals typically charge 39 percent more than their allowable operating costs. This "markup" varies across hospital types, somewhat reducing the precision of the findings.
- Omission of capital and direct medical education payments. This increases the apparent difference between charges and payments, especially because outlier cases are concentrated in teaching hospitals.

— PPS never intended to pay above-average costs incurred before the patient reached the cost or day outlier threshold.

— The marginal cost computed by Medicare appears to differ from the actual increments cost of care provided for care beyond the outlier threshold.

Guterman was unable to break down these four components precisely, but his analysis suggested very strongly that PPS did not protect hospitals fully against the marginal costs of outliers.

2. Outliers have very high charges and very high losses. Fear that hospitals would find a generous outlier policy profitable shaped the way the designers of PPS crafted the outlier payment system. However, outlier payments could probably be made more generous without giving hospitals an incentive either to keep patients past the outlier threshold or to recruit outlier cases.

3. Outliers appear to have more severe illness. The architects of PPS judged that marginal daily costs of outlier cases were 60 percent of average daily costs because they assumed that outliers who had passed the threshold were like typical patients nearing the end of a routine stay. The greater concentration of charges in special care and ancillary services suggests that outliers not only stay in the hospital longer but are sicker on average throughout their stay. This suggests that the marginal cost estimates may be significantly lower than actual marginal cost. For exceedingly ill patients, especially cost outliers, the marginal cost may be further underestimated because of the cross-subsidies inherent in per diem care charges, which we discussed in Section III. On the other hand, it is not possible to separate out those costs which occur after the outlier threshold is reached, so precise estimates are impractical.

4. PPS identifies stay and cost outliers in similar ways but pays them differently. The fact that stay and cost outlier threshholds identify patients with similar costs suggests equity. Because PPS pays a substantially larger fraction of

costs for cost outlier than for stay outliers may mean that a hospital loses a substantial amount on the day that a patient who would have been paid as a cost outlier crosses the stay outlier threshold and is paid as a stay outlier.

5. PPS paid less than planned under the outlier payment system. PPS paid about 53 percent of what it planned for outliers. In addition, the Department's target was based on the Federal portion, which will allow the target amount to increase substantially as the Federal portion increases. We have sought to expand payment under current law as much as we have discretion to do.

6. Outliers are Unevenly Distributed Across Hospital Types. This uneven distribution poses two problems. First, it means that even the law of large numbers will not protect classes of hospitals from outlier risks. Second, and of immediate practical importance, the adjustments for urban location and indirect teaching costs were written to take account of the tendency of these hospitals to have more than their share of outliers. When outlier payments were added on top of those adjustments, hospitals were effectively paid twice.

D. THE RELATION OF THE FINDINGS TO OUTLIER POLICY. Congress limited outlier payments to 5-6 percent of total payments. (Although Medicare actually paid much less than was authorized, this was the result of dramatic changes in length of stay which could not have been fully anticipated and of a decision not to apply outlier payments to the hospital-specific part of the payment.)

PPS legislation limited outlier payments to the "marginal" cost of services provided after the patient reached the "outlier threshold." By the time the threshold is crossed, charges are already substantially greater than Medicare payments.

Also, any case which could be paid as a "day outlier" would not be paid as a "cost outlier." This meant that Medicare's payment for a very expensive case which would have been a cost outlier decreased sharply when that patient stayed a day longer than the day outlier threshold and was therefore paid as a day outlier. HCFA is developing a revised payment approach for outliers which is meant to mitigate this problem.

Nevertheless, the legislative mandate set the volume of outlier payments and the fact that they were to pay marginal costs. The Guterman results invite us to address seven questions in light of the new data.

1. Does the Current Outlier Formula Focus Payments Effectively? If outlier payments are limited to a fixed percentage of payments, they should be focused in ways that best support the goals of the system:

— To protect hospitals against random risk, payments should be focused on exceptionally expensive cases occurring in smaller hospitals;

— To protect hospitals whose case mix has expensive features unmeasured by DRGs, payments should focus on those DRGs or conditions where the greatest differences in outlier costs can be shown across hospitals and hospital types; and

— To protect beneficiaries against discrimination, payments should focus on patients who might be identified at admission as having an unusually high risk of becoming outliers.

Each of these goals requires a specific strategy, and the current formula provides little focusing of payments. For all three goals outlier payments might provide more efficient protection if they paid an increasing fraction of costs as the gap between costs and DRG-based payment increases. This concept of increasing protection with increasing losses has been explored in some detail by Frank and Lave (1985) and McGuire et al. (1985) for psychiatric facilities and appears to have potential application to all outliers.

2. Can Cost Based Outlier Payments Be Increased Without Compromising PPS? Evidence as to the magnitude of hospital losses on outlier payments and the overall cost of these cases suggests perhaps an increase could be prudently made. Medicare

expects and appears to be getting cost-conscious behavior from hospitals in nonoutlier cases with far less stringent payment rules. A number of the Blue Cross plans and State Medicaid plans that have adopted DRGs for their Medicare programs, have used more liberal outlier policies than PPS.

3. Can a Severity-of-Illness System Replace the Current Outlier Payment System? One solution to outlier payment would be to replace it with a case mix system which deals accurately with outliers, something which the present DRG system is inherently unable to do. While we have indicated the limitations of severity measures with regard to many kinds of interhospital differences, such measures appear well suited to outlier cases, where severity of illness is clearly a major issue. For paying outliers, it might be reasonable to rely almost exclusively on a severity-based system where costs of collecting special data would be very small compared to the payment amounts involved.

However, outliers are likely to turn out to be largely patients who fail to respond to therapy and whose costs are, therefore, not accurately predictable even by excellent physiologic measures made at the time of admission. Further study in this area is clearly needed.

4. The 5-6 Percent Limit Medicare would need to spend more dollars than it did in the first year of PPS in order to provide substantial protection against the additional costs of outlier cases. Available data is not conclusive. The picture is especially complicated by the entry of New York and Massachusetts into PPS because these States are expected to have a very large volume of outliers and may reduce the availability of outlier payments for other States.

5. Effect on Other Adjustments. We have already pointed out that PPS originally paid twice for outliers in teaching hospitals and urban hospitals—it first adjusted payments in a way which reflected these hospitals' tendency to have more outliers and then paid extra for the individual outliers as they occurred. The

Omnibus Budget Reconciliation Act of 1986 corrected this problem for urban/rural differences by *adjusting* urban and rural rates separately for outlier payments, but the problem remains for teaching adjustments. Any change in outlier payment policy which increased payments to teaching hospitals would aggravate this problem unless corresponding adjustments were made to payments for indirect teaching costs.

6. Do "Marginal Costs" Define Appropriate Payments? This analysis highlights two problems with the marginal cost concept:

- Marginal costs are essentially undefinable unless substantial additional reporting requirements are imposed on hospitals. Even if such reporting requirements were imposed, estimating the relation of marginal costs to reported charges would remain very difficult. Recent data suggest that marginal costs are close to average costs and that the effort to define marginal costs more accurately may be of limited significance.
- True marginal cost is the hospital's actual cost for providing service, and paying true marginal cost means reverting, for outlier cases, to reimbursing full costs. Thus, once a patient had passed the outlier threshold, the hospital would be operating under the same inflationary incentives which led to PPS in the first place. This risk is significant because not all outlier costs result from appropriate care for which Medicare should pay.

These arguments suggest a need for careful research into the costs which hospitals actually incur in treating outliers. Based on such research, Medicare should be able to craft a policy which comes much closer to paying hospital costs without drifting back into the pre-PPS incentives. Such research should include careful examination of the marginal costs of very inexpensive cases to determine whether somewhat decreased payments for these cases might be approriate as an approach to funding higher payments for expensive outliers.

7. Why Do Patients Become Outliers? Guterman's study shows that the typical outlier patient is a disproportionate consumer of intensive care and ancillary services; in short, the average patient is sicker. But his analysis does not show the range of patient conditions and does not show whether a significant minority of outliers are seriously but not acutely ill patients who could be cared for in alternate settings; payment strategies for such patients might appropriately set lower payment rates. A second issue is the degree to which outliers are patients whose great cost is predictable at admission or whether they largely represent patients whose failure to respond to treatment is not predictable. Further research on these issues will be an important part of refining outlier payment policy.

E. A TYPOLOGY OF OUTLIERS. This analysis can be viewed from a different perspective by considering the relationship of why cases become outliers to the problems they present and the solutions which may be appropriate. Outlier cases can have a number of sources and awareness of this variety can clarify the problem of outlier payments. (The following language addresses cost, but the same logic applies to stay outliers.)

1. Qualitatively Different Cases which do not belong with other cases assigned to the DRG. These are typically relatively uncommon disorders which are too infrequent to have their own DRG but which are very different from other disorders in the same DRG. For example, acute necrosis of the liver, a catastrophic but uncommon disorder, is in DRG 205 along with a collection of liver diseases which are generally much milder and much less expensive to treat.

2. Patients needing unusual treatment which is very different in cost from treatments for other patients in the DRG. Radical surgery and bone marrow transplants illustrate this kind of outlier.

3. Patients with an unusually severe illness at the time of admission compared to other patients in the same DRG. Severity may be measured either directly by the

severity of the problem (a 90 percent burn is worse than a 70 percent burn) or by concurrent problems (an pneumonia is more serious in a patient with AIDS than in an otherwise healthy patient).

4. Patients who develop complications or who unpredictably fail to respond to treatment despite being apparently similar at admission to patients who have an uncomplicated course in the hospital.

5. Otherwise expensive patients who become outliers because they are treated in expensive facilities whether the result of their care is a better outcome or simply a higher bill.

This paradigm has implications for the kinds of problems which can be expected and for how improvements can be made. While groups 1, 2, and 3 may be expected to concentrate in specialized referral facilities, patients in group 4 should be randomly distributed across hospitals. We have already pointed out that patients in expensive facilities pose problems for which solutions are more distant. Thus, groups 1. 2, and 3 represent case mix problems while group 4 represents an insurance problem. To the extent that case mix problems are associated with hospital size, groups 1, 2, and 3 present special problems for large hospitals. On the other hand, group 4 presents special problems for small hospitals which are not protected from random events by a large volume of admissions. While improvements in the diagnosis coding and classification might be most useful for group 1 and improvements in procedure coding and classification might be most useful for group 2, admission severity measures would be more useful for group 3 and an insurance strategy focused on very high cost outliers would be most useful for group 4. While this typology does not provide solutions to any of the problems, it does show why a multifaceted strategy is necessary.

F. SUMMARY. Guterman's study makes clear that outlier payments are an area in which major steps can be taken to bring Medicare payments closer to costs.

Because the finite number of DRG groups makes coping with outliers an issue, outliers are an area in which DRG refinement is likely to be successful. Outlier payment reform, therefore, serves as an important complement to DRG refinement. Further research in this area has already started and should be completed in time for consideration for regulations effective October 1, 1987.

VII. SHOULD PPS PAY ACCORDING TO HOSPITAL SIZE?

We have noted repeatedly that the designers of PPS made a policy call when they decided that the increased costs of larger hospitals resulted from inefficiency and should not be recognized in the payment system. There was little empirical evidence either for or against their position. This section reviews the available evidence on this issue and examines some of the alternatives to directly paying more for care in larger hospitals.

A. EVIDENCE REGARDING HOSPITAL SIZE. Evidence that hospital size is associated with cost per case is well established. Every analysis of HCFA data and most other studies as well have confirmed the relationship. Section III pointed out that the effects of hospital size appear to primarily reflect the relatively low cost of hospitals smaller than a hundred beds rather than substantial differences among larger hospitals. These findings are not consistent with the view of the architects of prospective payment that increasing costs with increasing size represent inefficiency. If inefficiency resulted from large size, we would not expect it to taper off so abruptly at 100-250 beds. Rather, the most plausible explanation is that these smaller hospitals have diminished severity of illness or intensity of care.

Thus, the model of hospital size in the usual payment equation (Table 4.1A) is imprecise because it implies that costs continue to increase as hospital size increases. When a more accurate model is used (Table 4.1D), hospital size has little effect on the teaching adjustment (compare the coefficients for teaching in Tables 4.1B and 4.1D).

Thorpe et al.'s (1986) work suggests that DRG weight compression is associated with hospital size. Table 4.1D confirms this by showing that, when the equation is constructed to recognize the lower costs of small hospitals, the evidence of compression almost disappears (that is, the coefficient for case mix index becomes close to 1.0). Apparently, compression results in higher weights for cases found largely in smaller hospitals and lower weights for cases found largely in larger hospitals.

B. ALTERNATIVE APPROACHES. These technical findings suggest that even though paying hospitals according to size remains an undesirable policy, there are three strategies which might result in more equitable treatment of hospitals of different size without paying according to size.

1. Decompressing Weights. Compressed DRG weights increase the teaching adjustment. Decompressing the weights would result in somewhat larger payments to nonteaching hospitals with a complex case mix and somewhat smaller payments to teaching hospitals. This would be fairer to large nonteaching hospitals.

2. Increase Outlier Payments. Because outliers concentrate in large hospitals as well as in urban and teaching hospitals, outlier payments which came closer to costs will result in fairer payments to large hospitals.

3. Transfer Adjustments. (Bobula 1986) has suggested adjusting hospital payments by increasing payments to hospitals which receive transfers and decreasing payments to those making transfers. Since transfers overwhelmingly flow from smaller to larger hospitals, this approach would also be fairer to large hospitals.

VIII. HOSPITAL-LEVEL VARIABLES FOR CASE MIX ADJUSTMENT

In this section we examine the idea of adjusting payment at the hospital level and focus on some adjustors which may become available. There are two reasons why hospital-level adjustments may be preferred to patient-level adjustments:

--characterizing a hospital as efficient or inefficient or as having a simple or complex case mix may be easier than characterizing individual cases.

--if we wish to protect certain hospitals for policy reasons, this may be better achieved by hospital-level adjustments than by case mix adjustments.

The two most visible hospital-level adjustments in PPS are teaching status and urban/rural; we have already seen that PPS specifically does not recognize hospital size or allow the wage index to serve as a proxy for case mix beyond its direct fiscal effects on hospital costs. If the hospital can be characterized accurately, then it is possible to pay some or all cases at a rate which is related to costs actually incurred. There are two broad approaches to characterizing hospitals--historical evidence and analysis of current data.

PPS uses historical evidence for the hospital-specific portion of payments and for regional adjustments during the transition to national rates. Hospitals exempted or excluded from PPS are also paid on the basis of historical costs. Nevertheless, historical costs are most useful as a cushion to protect hospitals when the equity of a new rate system is uncertain or when transition may be too rapid.

Current data have at least two uses: characterizing certain tracer conditions to determine the efficiency of the hospital and characterizing certain aspects of case mix in order to make inferences about overall case mix.

A. TRACER CONDITIONS. A tracer is a condition which can be precisely and homogeneously defined so that different hospitals can be characterized by differences in the way they handle this standardized condition. Kessner (1973) pioneered this method in studying common nonhospital illnesses, but List et al., (1983) showed that variations in the severity of myocardial infarction could be controlled so as to use that disease as a tracer. Cromwell et al.'s (1985) work uses a variety of "tracers" which are generally less sharply defined. In principle, tracers can be used to define the relative efficiency of different hospitals, and this relative efficiency could then be as a basis for either hospital-specific or hospital type adjustments of payments. The difficulty with such methods is that cross-subsidization is so large for some conditions in some kinds of hospitals that costs for individual admissions cannot be accurately defined. This strategy will, therefore, require significant further development before it is fully satisfactory even as a research strategy.

B. TRANSFERS. We have already described work at HCFA by (Bobula 1986) that examines counting patient transfers as a tool for characterizing hospitals. They find that hospitals receiving more transfers have higher average costs than the typical hospital, even after correcting for case mix, wage index, location, and teaching activity. Conversely, hospitals sending more transfers have significantly lower than average costs per case. Measurement strategies of this kind may permit significantly greater accuracy in characterizing hospitals for payment purposes, but the policy implications of such measures present significant difficulties: in effect, a hospital might be paid quite heavily not to transfer patients. A related, and perhaps less manipulable, measure might be the distance which patients come to the hospital compared to nearby hospitals.

A very different way of characterizing hospitals is implicit in the finding tertiary and nontertiary hospitals have very different cost patterns. If this finding proves to be robust, patterns of costs may provide ways to characterize individual hospitals more accurately.

IX. THE ADEQUACY OF CURRENT CASE MIX ADJUSTMENT FOR NATIONAL RATES

The phased implementation of national PPS payment rates from a hospital-specific base provided both an opportunity to develop more refined case measures and time to determine the desirability and feasibility of moving to national rates. The foregoing analysis has made clear that major refinements of the case mix adjustment system are not ready, which places extra emphasis on the adequacy of the current system for national payment rates.

A. DECREASING MARGIN FOR ERROR IN PPS RATE SETTING. During the first years of PPS, most hospitals were protected from possible payment imprecision under national rates not only by the blend of regional and hospital-specific rates but also by high payments relative to historical costs and substantial gains in efficiency. In the immediate future, hospitals will be pressed not only by the removal of these protections but also by an increasingly competitive health care environment, problems in financing uncompensated care, and a Federal budget which is increasingly constrained by "Gramm-Rudman-Hollings" (the Balanced Budget and Emergency Deficit Control Act of 1985: P.L. 99-177). Hospitals will have fewer and fewer alternative sources of income to which they can turn if Medicare's payments are too low. On the other hand, Medicare currently provides average margins in excess of 12 percent, which leaves substantial protection against any error.

B. THE LIMITATIONS OF THE CURRENT CASE MIX ADJUSTMENT SYSTEM. Differences in hospital costs appear to stem from a variety of factors which are not reflected in PPS. We have reviewed studies which suggest that there are significant differences in:

--severity of illness within DRGs, not only among patients but also among hospitals,

--the way hospitalization is used, not only among patients and hospitals but also across States and perhaps regions,

--local standards for appropriate intensity of care, among hospitals, types of hospitals, and States, and

--costs, resulting from all of these factors and also from patterns of referral, kinds of admissions, and specialization in certain kinds of disease.

C. STRATEGY FOR NATIONAL RATES. Because we do not have satisfactory measures of severity of illness or satisfactory ways to assess what intensity of care is medically appropriate, we cannot resolve the question of which differences between costs and payments are appropriate. Clearly, perfect correlation between costs and payments would be equivalent to cost-based reimbursement; because there still appears to be substantial room for increasing efficiency in the hospital industry, payments which perfectly reflect costs would not be a desirable goal. On the other hand, we cannot reach a technical judgment whether the degree of match between costs and payments which we have found, and which we project to occur under national rates, is the correct one. The difference between costs and payments may reflect correctable inefficiencies for which Medicare should not pay; if this is true, then the high operating margins realized by some hospitals are useful incentives to other hospitals to become more efficient. On the other hand, the differences between costs and payments may mean that hospitals with severely ill patients and medically appropriate intense care are being penalized by a system which does not fully measure their special responsibilities. Current technology does not allow us to determine which of these two scenarios is most correct.

Under these circumstances, close monitoring of the impact of PPS on providers and patients becomes especially important, and close cooperation between Congress and the Administration in charting the future course of the program will be essential. The refinements of case mix measurement which could improve payment of severity and complexity are also the research tools necessary to determine whether PPS payments are equitable; further refinement of case mix adjustment methods is an integral part of the future of PPS. The next chapter describes an agenda for work in this area.

D. SUMMARY. We have examined the reasons why intensity of care differs across hospitals. We found that the best current estimates indicate that the impact of PPS has been somewhat different than that which was predicted and raised some question that the current case mix adjustment system is sufficiently precise for national rates. We identified the outlier payment system as the area in which greatest progress can be made most rapidly to bring PPS payments into closer alignment with hospital costs. Finally, we suggested ways in which case mix measurement could make greater use of hospital-level adjustments without compromising the basic incentives of PPS.

Table 4.1

PREDICTORS OF HOSPITAL AVERAGE COST PER CASE
Percentage Increase for Every 1.0 Percent Increase in Variable

	1981 Costs	1984 Costs	
A: Basic Model			
Case Mix Index	1.014	0.954	
Teaching*	0.579	0.583	
Wage Index	1.019	0.959	
Large SMSA	11.60	9.91	)percent
Medium SMSA	2.76	3.78	)higher
Small SMSA	0.10	2.54	)than rural
Number of Beds	0.119	0.086	
R^2	0.725	0.702	
B: Without Hospital Size (beds)			
Case Mix Index	1.464	1.295	
Teaching*	0.845	0.766	
Wage Index	0.980	0.863	
Large SMSA	23.18	19.97	)percent
Medium SMSA	11.96	11.34	)higher
Small SMSA	9.07	9.35	)than rural
R^2	0.692	0.682	
C: Coefficients Reflecting Actual 1984 Payment Procedures			
Case Mix Index	1.000[1]	1.000[1]	
Teaching*	1.158[1]	1.158[1]	
Wage Index	0.7915[1]	0.7915[1]	
Urban	25.34	18.70	percent higher than rural
R^2	0.666	0.669	

[1] Value mandated by law or regulation.

* Teaching is defined as the ratio of interns and residents to hospital beds.

1981 (N=4971): hospitals with >49 Medicare discharges in 1981 and any 1984 discharges.

1984 (N=5349): hospitals with >49 Medicare discharges and with available Cost Reports.

(continued)

Table 4.1
(continued)

PREDICTORS OF HOSPITAL AVERAGE COST PER CASE
Percentage increase for Every 1.0 Percent Increase in Variable

	1981 Costs	1984 Costs
D: With Hospital Bed Size Groups		
Case Mix Index	1.097	1.003
Teaching*	0.742	0.702
Wage Index	1.001	0.943
Large SMSA	13.21	10.57)) percent
Medium SMSA	3.91	4.42)) higher
Small SMSA	1.04	2.87)) than rural
51-100 Beds	14.11	9.10) percent
101-250 Beds	25.24	18.27) higher than
251-500 Beds	29.55	21.07) 1-50 bed
More than 500 beds	32.81	21.15) hospitals
R^2	0.725	0.702
E:		
Case Mix Index	2.851	2.298
R^2	0.401	0.423
F:		
Case Mix Index	1.469	1.376
Teaching*	0.899	0.846
Wage Index	1.436	1.191
R^2	0.670	0.669

* Teaching is defined as the ratio of interns and residents to hospital beds.

1981 (N=4971): hospitals with >49 Medicare discharges in 1981 and any 1984 discharges.

1984 (N=5349): hospitals with >49 Medicare discharges and with available Cost Reports.

Source: Medicare Cost Reports.

Table 4.2

PREDICTORS OF HOSPITAL AVERAGE COST PER MEDICARE CASE IN 1984

Variable	Effect[1]	Cumulative[2] R^2
Wage Index	0.971	0.487
Case Mix Index	1.275	0.641
Teaching Activity	0.686	0.661
Urban location	11.068	0.672
Disproportionate share[3]	0.719	0.674
Including Outlier Payments		
Wage Index	0.891	0.487
Case Mix Index	1.253	0.641
Teaching Activity	0.336	0.661
Urban location	6.716	0.672
Disproportionate share[3]	0.133	0.674
Outlier payments[4]	2.571	0.737
Controlling for Hospital Size		
Wage Index	1.031	0.482
Case Mix Index	0.951	0.641
Teaching Activity	0.618	0.661
Urban location	3.821	0.672
Disproportionate share[3]	0.497	0.674
50-99 beds	9.771	)
100-149 beds	17.626	)
150-199 beds	18.338	) 0.696
200-399 beds	22.308	)
more than 399 beds	25.028	)

[1] Effects correspond to coefficients in a log-linear regression model which includes all of the variables. For continuous variables (wage index, case mix index, teaching activity, disproportionate share, and outlier payments) the effect is the percent increase in cost corresponding to a 1.0 percent increase in the variable. For dummy variables (urban location and bed size) the effect is the percent increase in cost per case if hospital is urban or if it has the indicated number of beds, compared to a rural hospital or to one with 1-49 beds.

[2] Cumulative R^2 is the portion of the total variance in hospital average cost per case which is accounted for in a stepwise regression by variables down to the given value. All hospital size dummies are entered together.

[3] The disproportionate share variable is the percentage increase in payments to the hospital which is defined by legislation; the coefficient indicates the percent increase in cost which corresponds to a 1.0 percent increase in payments required by law.

[4] Outlier payments are the fraction of payments for the hospital made to the hospital estimated from casemix during the first year of PPS and payment rules for Federal Fiscal Year 1986 but assuming full implementation of PPS.

Source: Medicare Cost Reports. N=5350 hospitals.

Table 4.3

DEATHS DURING ADMISSION AND WITHIN 30 DAYS AFTER DISCHARGE
5 Percent Sample of 1984 Medicare Discharges

	Total Adms	Deaths Within 30 Days	Deaths During Admission	Difference	Average Length of Stay
Oregon	5995	8.4%	4 9%	3.5%	6.51
Utah	1897	8.8%	5.6%	3.2%	6.64
Wyoming	725	7.9%	4.8%	3.0%	7.52
Washington	8292	8.0%	5.3%	2.7%	6.86
Montana	1802	8.2%	5.7%	2.6%	7.00
Iowa	7573	7.4%	5.0%	2.4%	8.12
Kansas	6466	7.5%	5.1%	2.4%	8.67
Minnesota	9503	7.6%	5.2%	2.4%	7.69
Nevada	1691	8.6%	6.3%	2.4%	7.81
Arkansas	7385	8.3%	5.9%	2.4%	7.30
S Dakota	2214	6.0%	3.7%	2.2%	7.45
Wisconsin	11278	7.4%	5.2%	2.2%	8.35
Arizona	5946	7.7%	5.4%	2.2%	7.92
Oklahoma	7779	8.5%	6.3%	2.2%	7.75
N Mexico	2246	6.9%	4.7%	2.2%	7.32
Colorado	4692	7.1%	4.9%	2.2%	7.84
Nebraska	4491	7.7%	5.6%	2.1%	7.94
Georgia	13633	7.4%	5.4%	2.0%	7.60
Missouri	15294	7.6%	5.5%	2.0%	8.94
Kentucky	9256	7.5%	5.5%	2.0%	8.46
Texas	33071	7.5%	5.5%	2.0%	8.14
N Dakota	2260	6.3%	4.4%	1.9%	7.79
California	44630	7.6%	5.7%	1.9%	7.71
Louisianna	9814	7.4%	5.5%	1.9%	7.85
Idaho	1807	6.8%	4.9%	1.8%	6.50
Tennessee	14757	7.0%	5.2%	1.8%	8.65
Indiana	12871	7.6%	5.8%	1.8%	8.89
Vermont	784	7.9%	6.1%	1.8%	8.93
Alablam	10603	6.6%	4.8%	1.8%	8.15
Mississippi	7741	6.8%	5.1%	1.8%	7.76
Ohio	21572	7.6%	5.9%	1.8%	9.32
Delaware	1201	8.7%	7.0%	1.7%	10.41
Florida	34209	7.3%	5.6%	1.7%	8.42
Hawaii	1156	8.0%	6.3%	1.6%	8.74
N Carolina	12150	8.2%	6.6%	1.6%	8.83
Illinois	27084	7.5%	5.9%	1.6%	9.72
S Carolina	6282	7.2%	5.8%	1.5%	8.84
Penna.	34868	7.3%	5.9%	1.4%	10.03
Michigan	19938	7.3%	5.9%	1.4%	9.31
New Hamp.	1363	6.9%	5.5%	1.4%	8.61
Virginia	10678	7.5%	6.1%	1.3%	9.28
Maine	3076	7.3%	6.0%	1.3%	8.44

(continued on next page)

TABLE 4.3
(Continued)

DEATHS DURING ADMISSION AND WITHIN 30 DAYS AFTER DISCHARGE
5 Percent Sample of 1984 Medicare Discharges

	TOTAL ADMS	DEATHS WITHIN 30 DAYS	DEATHS DURING ADMISSION	DIFFERENCE	AVERAGE LENGTH OF STAY
W. Virginia	5769	6.7%	5.4%	1.2%	8.17
Rhode Is.	2193	7.7%	6.6%	1.1%	10.33
Connecticut	6392	8.1%	7.1%	1.0%	9.84
Maryland	6107	7.4%	6.7%	0.8%	10.33
Dist. Col.	1293	6.7%	6.0%	0.7%	12.20
Mass.	14560	7.2%	6.7%	0.5%	11.58
Alaska	226	4.0%	3.5%	0.4%	7.96
New Jersey	15837	7.7%	7.6%	0.1%	12.11
New York	37973	7.8%	8.6%	-0.8%	13.82
U.S.	530423	7.5%	5.9%	1.6%	9.11

Source: Eggers P. Unpublished Data.

Table 4.4

Hospital Admission and Surgery Rates
Selected DRG Clusters and Procedures
Per 1000 Medicare Beneficiaries
1982-1983 Medicare Data

	Michigan	New Jersey	North Carolina	Washington
Cerebrovascular Disease				
Admissions	15.6	16.3	16.1	12.5
Extracranial Vascular Procs	1.0	0.8	0.9	2.0
Coronary Artery Disease				
Admissions	19.7	16.2	22.5	13.5
Coronary Artery Bypass Grafts	0.9	0.4	0.8	2.2
Cardiac Catheterizations	2.2	2.0	2.8	3.1
Pneumonia				
Admissions	7.5	6.8	9.5	6.9
Prostate Disease				
Admissions	7.4	5.9	7.0	6.3
Prostatectomies	5.6	4.6	5.0	5.5
Admissions without Prostatectomy	1.8	1.3	2.0	0.8

Source: Mitchell et al. 1986

Table 4.5

Death Rate Per 100 Admissions By State By DRG Within DRG Cluster 1982-3 Medicare Data

Cerebrovascular Disease	Michigan	New Jersey	North Carolina	Washington
DRG 5	1.2	4.1	1.7	2.9
DRG 14	17.1	20.2	16.3	15.3
DRG 15	1.1	2.0	0.9	1.1
Coronary Artery Disease				
DRG 106	7.4	9.1	5.1	4.2
DRG 125	2.9	10.3	4.3	4.0
DRG 132	6.0	7.5	6.4	5.9
DRG 140	0.9	1.1	0.6	2.2
Pneumonia				
DRG 79	22.8	25.9	15.2	16.8
DRG 80	16.3	11.8	2.7	4.3
DRG 89	13.1	15.2	12.6	9.1
DRG 90	5.9	4.3	1.0	4.2
Prostate Disease				
DRG 306	1.1	1.4	1.5	0.8
DRG 336	0.5	0.9	0.6	0.2
DRG 346	15.4	22.2	15.4	15.1
DRG 348	1.2	2.6	1.4	0.0

Source: Mitchell et al. 1986

Table 4.6

Effect of Decompressing* DRG Weights With and Without Inclusion of Hospital Bedsize Case Mix Index Coefficient Constrained to Reflect Payment Rules Rand 1981 Synthetic Costs

MODEL A	"Compressed"	"Decompressed"
Case Mix Index	10.00	10.00
Teaching	7.85	6.89
Wage Index	9.03	8.63
SMSA	15.65	13.60 percent higher than rural
R^2	0.590	0.596

MODEL B	"Compressed"	"Decompressed"
Case Mix Index	10.00	10.00
Teaching	3.55	3.52
Wage Index	10.26	10.29
SMSA	8.35	2.93
Hospital Beds	9.45	9.06
Regional Differences	-3.3 to +10.7	-2.9 to +10.9 percent compared to pacific region
R^2	0.633	0.633

*See text.

Source: Thorpe et al 1986

Table 4.7

Average Operating Margins on Medicare Cases by Class of Hospital First Hospital Fiscal Years on Prospective Payment

Hospital Type	Number of Hospitals	Operating Margin %
National Average	4544	16.4
Urban	2217	18.2
1-99 beds	545	14.2
100-404 beds	1347	16.9
405-684	277	20.0
more than 684 beds	48	23.0
Rural	2327	8.6
1-49 beds	1074	6.0
1-99 beds	1795	8.3
100-169 beds	357	8.4
more than 169 beds	175	9.3
Teaching status		
Nonteaching	3762	13.4
Light (<.25 res/bed)	660	19.4
Heavy (>.25 res/bed)	122	23.3
Disproportionate share		
Non-DSH	3636	15.6
Urban DSH >100 beds	670	18.7
Urban DSH <100 beds	59	11.7
Rural DSH	179	8.7
Sole Community/Rural Referral		
Non-SCH/RRC	4082	16.9
Sole Community	289	7.0
Rural Referral Center	143	10.2
Both SCH and RRC	16	9.9
Ownership		
Voluntary	2700	17.4
Proprietary	649	14.3
Government	1178	13.2

Table 4.8

Medicare Operating Margin By Urban/Rural Location and Number of Beds

Actual First Year of PPS and Projected Experience under Full National Rates

		Percent of Hospitals									Number of Hosps
		Loss				Profit					
		>20%	20-10%	10-5%	5-0%	0-5%	5-10%	10-20%	20-30%	>30%	
Urban <100 beds	PPS yr 1	4	4	5	6	11	11	28	19	13	545
	National rates	6	5	5	6	8	7	14	10	40	675
Urban 100-404 beds	PPS yr 1	1	1	1	4	8	13	35	25	12	1347
	National rates	3	6	5	9	10	10	20	14	21	1649
Urban 405-684 beds	PPS yr 1	0	0	0	3	5	13	32	29	18	277
	National rates	0	8	5	8	11	11	21	17	20	332
Urban >684 beds	PPS yr 1	0	0	0	0	1	3	15	16	13	48
	National rates	0	10	7	3	11	11	21	17	20	70
Rural <100 beds	PPS yr 1	6	8	7	11	12	12	26	12	7	1795
	National rates	4	7	5	6	8	9	18	14	30	2029
Rural 100-169 beds	PPS yr 1	1	4	7	10	17	17	31	11	3	357
	National	4	10	6	11	11	8	20	15	13	397
Rural >169 beds	PPS yr 1	1	3	4	10	16	22	31	10	3	175
	National	4	7	5	7	11	10	26	14	17	197
All hosps	PPS yr 1	3.0	4.4	4.5	7.6	10.8	12.6	29.9	17.7	9.5	4544
	National	3.7	6.9	5.2	7.5	9.0	9.4	18.6	13.8	26.0	5349

Table 4.9

Medicare Operating Margins By Urban/Rural Location and Teaching Status

Actual First Year of PPS and Projected Experience under Full National Rates

		Percent of Hospitals									Number of Hosps
		Loss				Profit					
		>20%	20-10%	10-5%	5-0%	0-5%	5-10%	10-20%	20-30%	>30%	
Urban non-teaching	PPS yr 1	2	2	3	5	10	13	32	22	11	1491
	National	4	6	6	8	10	9	18	13	27	1823
Urban light teaching	PPS yr 1	0	1	1	3	6	12	27	37	15	608
	National	2	7	6	9	9	12	20	13	21	744
Urban heavy teaching	PPS yr 1	0	1	0	3	4	3	27	30	31	118
	National	2	6	5	6	8	7	21	12	33	159
Rural non-teaching	PPS yr 1	5	7	7	11	13	13	27	12	6	2271
	National	4	7	5	7	9	9	19	14	26	2556
Rural light teaching	PPS yr 1	2	4	2	2	17	15	37	13	8	52
	National	3	10	8	2	8	5	27	20	17	60
Rural heavy teaching	PPS yr 1	0	0	0	0	0	25	0	25	50	4
	National	0	14	0	0	0	14	0	14	57	7
All hosps	PPS yr 1	3.0	4.4	4.5	7.6	10.8	12.6	29.9	17.7	9.5	4544
	National	3.7	6.9	5.2	7.5	9.0	9.4	18.6	13.8	26.0	5349

Table 4.10

Predictors of Hospital Medicare Operating Margins with and without control for outlier payments

Variable	PPS Year 1[1]	National Rates[1]
Wage Index	-0.054**	-0.244****
Case Mix Index	0.370****	0.032
Teaching Intensity	0.107**	0.194****
Urban	5.232****	10.950****
Disproportionate share[2]	0.140*	0.492****
Less than 50 beds	0.00	0.00
50-99 Beds	2.498****	-6.838****
100-149 beds	1.113	-10.969****
150-199 beds	0.476	-9.875****
200-399 beds	0.501	-12.380****
More than 399 beds	0.728	-13.552****
Adjusted R^2	0.151	0.057
Wage Index	-0.043**	-0.201****
Case Mix Index	0.384****	-0.027
Teaching Intensity	0.119***	0.287****
Urban	5.364****	10.191****
Disproportionate share[2]	0.235**	0.644****
Less than 50 beds	0.00	0.00
50-99 Beds	2.882****	-5.679****
100-149 beds	1.633*	-9.197****
150-199 beds	1.226	-7.645****
200-399 beds	1.362*	-9.279****
More than 399 beds	0.541	-8.713****
Outliers	-0.507****	-0.973****
Adjusted R^2	0.158	0.084
Number of Hospitals	4543	5313

*= p <.05; ** = p<.01; *** = p<.001; **** = p<.0001

[1] Effects correspond to coefficients in a log-linear regression model which includes all of the variables. For continuous variables (wage index, case mix index, teaching activity, disproportionate share, and outlier payments) the effect is the percent increase in cost corresponding to a 1.0 percent increase in the variable. For dummy variables (urban location and bed size) the effect is the percent increase in cost per case if hospital is urban or if it has the indicated number of beds, compared to a rural hospital or to one with 1-49 beds.

[2] The disproportionate share variable is defined by legislation; the coefficient indicates the percent increase in margin which corresponds to a 1.0 percent increase in payments required by law.

[3] The coefficient for outliers reflects the change in operating margin corresponding to a 1.0 percent increase in the percentage of a hospital's payments which is received from outliers.

Source: Medicare Cost Reports.

Table 4.11

Variations in Hospital Medicare Operating Margins Across Regions Controlling for other hospital characteristics

Variable	PPS Year 1[1]	National Rates[1]
New England	0.00	0.00
Mid Atlantic	+5.43****	+7.71****
South Atlantic ex FL	+1.02	+9.55****
South Atlantic: FL	-0.10	-0.59
East N Cent: IL & MI	-4.89****	-1.41
East N Cent: IN, OH, WI	3.63***	6.56****
East S Cent: MS	3.43*	14.48****
East S Cent: KY, TN, AL	-0.16	10.80****
West N Cent: NB, ND, SD	2.38	16.36****
West N Cent: IO, KA, MO, MN	1.47	9.07****
West S Cent: TX	1.08	7.71****
West S Cent: AR, LA, OK	1.09	11.11****
Mountain: UT	2.42	14.53****
Mountain: AZ, NM, NV	-1.81	-1.03
Mountain: ID, MT, CO, WY	3.52**	10.62****
Pacific: CA	1.77	3.93*
Pacific: AK, HI	-2.81	7.37
Pacific: OR, WA	3.14*	16.35****
Wage Index	-0.084**	-0.147***
Case Mix Index	0.382****	-0.011
Teaching Intensity	0.097**	0.182***
Urban	6.097****	9.359****
Disproportionate share[2]	0.173*	0.479****
Less than 50 beds	0.00	0.00
50-99 Beds	2.498****	-5.887****
100-149 beds	1.113	-10.841****
150-199 beds	0.476	-9.134****
200-399 beds	0.501	-12.012****
More than 399 beds	0.728	-13.669****
Adjusted R^2	0.163	0.091

*= p <.05; ** = p<.01; *** = p<.001; **** = p<.0001

For notes see Table 4.10. Coefficients for regions are relative increase or decrease in operating margin compared to a otherwise similar hospital in the New England region.

Source: Medicare Cost Reports.

TABLE 4.12

Correlations in 50 States of Costs of 8 Highest Dollar Volume DRGs
1984 PPS Discharges
Rand Synthetic Costs

DRG	14	209	89	468	148	210	140
127 - Heart Failure and Shock	.64	.51	.54	.41	.40	.46	.61
14 - Cerebrovasc Disorders X TIA		.58	.57	.30	.22	.23	.43
209 - Major Joint Procedures			.74	.60	.45	.34	.58
89 - Simple Pneumonia Age > 70				.39	.26	.34	.61
468 - Unrelated OR Procedure					.02	.70	.42
148 - Major Small & Large Bowel Proc Age > 70						-.08	.12
210 - Hip or Femur Proc X Major Jt, Age > 70							.54

Costs are Rand synthetic costs for 1984.

Source: Cretin et al. 1986

Table 4.13

Characteristics of Medicare Outlier Cases Cases Paid Under PPS in Federal Fiscal Year 1984

	Type of Outlier				
Variable	Length of Stay	Cost	All	Non-Outliers	All PPS Cases
PPS Discharges	1.76%	1.21%	2.96%	97.04%	100.00%
Urban Hospitals	2.07	1.45	3.52	96.48	100.00
Rural Hospitals	0.98	0.60	1.58	98.42	100.00
Teaching Hospitals	2.58	1.24	3.82	96.18	100.00
Nonteaching Hospitals	1.34	1.19	2.53	97.47	100.00
Hospital Days	9.17	3.34	12.51	87.49	100.00
Total Charges	10.49	7.17	17.66	82.36	100.00
Est. Operating Costs	10.38	6.80	17.19	82.81	100.00
Medicare Payments	3.35	3.00	6.35	93.65	100.00
Charges less Payments	34.71%	21.32	56.02	43.98	100.00
Special care charges (percent of total)	10.10	14.20	11.76	6.36	7.32
Ancillary charges (percent of total)	62.12	63.89	62.83	55.68	56.48
Length of Stay	40.9	21.7	33.1	7.1	7.8
Total Charge	$ 26,151	$26,012	$26,094	$ 3,714	$ 4,377
Charge per Day	$ 639	$1,200	$788	$ 526	$ 559
Est. Operating Cost	$ 17,116	$16,385	$16,819	$2,473	$2,898
Medicare payment	$ 6,457	$8,399	$7,248	$3,262	$3,380
Est. Operating Margin	$-10,658	$-7,989	$-9,573	$+791	$+484

Source: Stuart Guterman, Health Care Financing Administration, unpublished data.

Table 4.14
DISTRIBUTION OF OUTLIER CASES BY HOSPITAL GROUP
PPS Discharges in Federal Fiscal Year 1984

	Outliers as a Percentage of PPS Cases		
Hospital Group	Length of Stay Outliers	Cost Outliers	Total Outliers
All Hospitals	1.76%	1.21%	2.96%
Urban	2.07	1.45	3.53
o <100 beds	0.93	0.92	1.85
o 100-404 beds	1.78	1.59	3.37
o 405-684 beds	2.59	1.39	3.98
o 685+ beds	3.04	1.23	4.27
Rural	0.98	0.60	1.58
o <100 beds	0.64	0.40	1.04
o 100-169 beds	0.97	0.68	1.65
o 170+ beds	1.61	0.87	2.47
New England	2.84	0.59	3.42
Mid-Atlantic	2.05	1.38	3.44
South Atlantic	1.90	1.45	3.35
East North Central	2.10	0.84	2.94
East South Central	1.72	1.22	2.95
West North Central	1.48	0.91	2.39
West South Central	1.32	1.02	2.34
Mountain	1.16	1.40	2.56
Pacific	1.33	1.92	3.25
Major Teaching	3.39	1.15	4.54
Other Teaching	2.46	1.26	3.71
Non-Teaching	1.34	1.19	2.53
Not-for-Profit	1.89	1.15	3.04
Proprietary	1.33	2.20	3.53
Government	1.47	0.86	2.33

Source: Health Care Financing Administration, Office of Research and Demonstrations.

CHAPTER 5

RESEARCH FOR IMPROVING CASE MIX MEASUREMENT UNDER PPS

This chapter describes work in progress on refining case mix measurement and lays out refinements which may become possible in each of the three major areas we have examined:

— improving case mix measurement at the individual case level,

-- improving the payment of outlier cases, and

— improving hospital-level adjustments.

I. IMPROVING MEASUREMENT AT THE INDIVIDUAL CASE LEVEL

We have pointed out that there is no single tested, feasible, off-the-shelf solution to measuring patient treatment needs more accurately. Over the next 3 years, however, a series of refinements show promise of becoming feasible:

A. REFINEMENTS OF THE USE OF SECONDARY DIAGNOSES IN THE DRG SYSTEM. Work nearing completion at Systemetrics under Jon Conklin and work just beginning at Yale under Robert Fetter promise substantial improvements in the use of secondary diagnoses to identify severely ill patients within the DRG system. HCFA will be evaluating the Systemetrics work over the next year and expects first results from the Yale work in late 1987.

B. EXTEND RESEARCH ON USING PHYSIOLOGIC MEASUREMENTS TO REFINE DRGS. HCFA is already sponsoring a number of projects in this area:

1. Health Data Institute (Paul Gertman Principal Investigator). This is a multidimensional study of the use of physiologic measures and other data to predict costs, including an integration of inpatient and outpatient data from several sources.

2. Tulane School of Public Health (Farouk Alemi Principal Investigator). This is a study of seven different instruments to predict the cost of care for patients with myocardial infarction in two hospitals in the greater New Orleans area.

3. The Potential of MEDISGRPS for DRG Refinement (Mark Moskowitz Principal Investigator). This study is under the auspices of the Brandeis Health Policy Center. It will examine existing data from hospitals subscribing to the MEDISGRPS system to determine how much payment accuracy MEDISGRPS could add to DRGs.

In addition, an interesting study is not funded by HCFA:

4. Alternative Severity Measures (William Thomas and Marie Ashcraft Principal Investigators). This study is funded by a consortium of university hospitals and the State of Michigan. It compares APACHE, Computerized Disease Staging, Clinical Disease Staging, Patient Management Categories, and DRGs for reliability, clinical validity, and ability to predict cost and length of stay. Preliminary results do not indicate that any available system is a major improvement over DRGs.

D. COMPARATIVE EVALUATION OF PHYSIOLOGIC MEASURES OF SEVERITY OF ILLNESS AS REFINEMENTS TO DRGs. An evaluation of these measures would compare at a minimum, APACHE, Computerized Severity Index, and MEDISGRPS. Ideally, the study would also test a number of techniques identified in the work described in Section B. The study would use a nationally representative data base and would measure vulnerability to manipulation, reliability, and costs of data collection. The study would assess the usefulness of these methods both for payment research and for quality of care assessment (Chapter 3.VII discusses design and methodology in somewhat more detail). HCFA is currently exploring the design of this study as well as its feasibility.

E. ANALYZE DATA COLLECTED IN PRO QUALITY ASSESSMENT EFFORTS. Data collected by PROs to assess quality of care will generally also bear on severity of patient illness. HCFA is currently seeking ways to maximize the usefulness of PRO data for studies of the relation between costs and both severity and intensity. In addition, PRO data may be usable to differentiate reimbursable and inappropriate outlier cases.

F. INTERNAL HCFA RESEARCH. HCFA will continue to pursue intramural studies of the relation between beneficiary characteristics and costs at the individual DRG level, following up initial findings that differences are small when DRGS are aggregated but quite significant for individual DRGs.

II. IMPROVE PAYMENT OF OUTLIER CASES

HCFA is analyzing Medicare cost reports for the first year of PPS. With this information it will be possible to estimate the impact of the current outlier payment strategy more precisely, to project the impact of paying larger amounts under the present general structure, and to assess what changes might be needed in the adjustments for teaching and urban/rural location to reflect the fact that urban hospitals and teaching hospitals get the majority of outlier payments. This work includes careful study of the merits of focusing outlier payments on certain DRGs and of the applicability of physiologic measures to outliers. Studies in this area will include securing available information and experience from Medicaid, Blue Cross, and other plan which are using DRGs with outlier payment policies different from those of Medicare.

The Rand-UCLA Health Policy Research Center is currently analyzing outlier payments with special focus on policy alternatives which do not require legislative changes.

III. IMPROVING ADJUSTMENTS AT THE HOSPITAL LEVEL

The major project in this area is:

Boston University School of Medicine (Lisa Iezzoni Principal Investigator). This study will examine the differences in patient populations between teaching and nonteaching hospitals in the Boston area.

In addition, internal HCFA research will be addressed at three issues:

- Can hospitals be characterized effectively by their sending of and receipt of transfer patients?
- Can HCFA use differences in cost patterns between similar DRGs in teaching and nonteaching hospitals to characterize teaching activity more accurately than the resident/bed ratio?
- Does the distance patients travel to a hospital, compared to other hospitals in the same area, provide an index of referral center status?

References

AHA: Memorandum from Edmund B. Rice to Jack W. Owen. Washington, D.C. American Hospital Association, November 25, 1985.

Altman, S.: Testimony before the Health Subcommittee, Committee on Finance, U.S. Senate. July 29, 1985.

Arthur Young Co.: A Study of Financing Graduate Medical Education: Data Collection. Competitve contract No. HHS-100-80-0155. Washington D.C. Arthur Young Co., January 1983, updated September 1983.

Averill, R.F.: HCFA's Process for Refining the DRG Classification System. New Haven Conn. Health Systems International, 1986. *

Averill, R.F., Mullin, R.L., and Elias, E.D.: The Revised ICD-9-CM Diagnosis Related Groups. New Haven, Conn. Health Systems International, Undated documentation to the May 1983 software.

Bedell, S.E., and Delbanco, T.L.: Choices About Cardiopulmonary Resuscitation in the Hospital. N Eng J Med 310(17): 1089-1093. April 26, 1984.

Beebe, K., Lawrence, T., and Lintzeris, G.: Medicare Admissions and Length of Stay for Short-Stay Hospitals. Health Care Financing Review 7(1): 117-119. Fall 1985.

Bobula, J.D.: The Impact of Hospital to Hospital Transfers on Hospital Costs. Working Paper. Baltimore, Md. Health Care Financing Administration, March, 1986.

Bobula, J.D.: Unpublished data. Baltimore, Md. Health Care Financing Administration, 1986.

Brewster, A.C., Bradbury, R.C., Chae, Y.M., et al.: DRG Variance Analysis Using the Medical Illness Severity Grouping System (MEDISGRPS): Determining the Relative Impact of Admission Severity, Medical Care Effectiveness and Treatment Efficiency on Hospital Revenue under Prospective Payment Methods. Prepublication draft. West Borough, Mass. MediQual, July 1984a.

Brewster, A.C., Bradbury, R.C., and Jacobs, C.M.: Measuring the effect of illness severity on revenue under DRGs. Healthcare Financial Management 52-60, July 1985a.

Brewster, A.C., Karlin, B.G., Hyde, L.A., et al.: Medical Illness Severity Grouping System (MEDISGRPS): A Clinically-Based Approach to Classifying Hospital Patients at Admission. Mimeo. West Borough, Mass. MediQual, Undated.

---

*References marked with an asterisk were commissioned for this report.

Brewster, A.C., Karlin, B.G., Hyde, L.A., et al.: Medical Illness Severity Grouping System (MEDISGRPS): A Clinically-Based Approach to Classifying Hospital Patients at Admission. Prepublication draft. West Borough, Mass. MediQual, November 9, 1984b.

Brewster, A.C., Karlin, B.G., Hyde, L.A., et al.: Medical Illness Severity Grouping System (Medisgrps**): A Clinically-Based Approach to Classifying Hospital Patients at Admission. Inquiry, in press. Prepublication Draft. West Borough, Mass. MediQual, July 29, 1985b.

Brewster, A.C., Bradbury, R.C., Hyde, L.A., et al.: A Computerized Clinical Information System for Hospital Quality Control. Prepublication draft. West Borough, Mass. MediQual, November 20, 1984c.

Butler, P.W., and Bentley, J.D.: The Disease Staging Case Mix of a Sample of Teaching Hospitals: A Technical Report. Washington D.C. Association of Medical Colleges, 1982.

Calore, K.A.: Patient Management Categories and Disease Staging: Can They Improve DRGs? Health Care Financing Administration Co-operative Agreement No. 18-C-98526/1-02. Waltham, Mass. Health Policy Research Consortium - Cooperative Research Center, December 27, 1985. *

Carter, G., and Ginsberg, P.: Changes in Coding Practices Under PPS: Are They Self-Limiting? Health Care Financing Administration Co-operative Agreement No. 18-C-98489/9-01. Santa Monica, Calif. Rand Corp., 1986. *

Chassin, M.R.: Variations in Hospital Length of Stay: Their Relationship to Health Outcomes. Health Technology Case Study 24, OTA-HCS-23 Washington D.C. U.S. Congress, Office of Technology Assessment, August 1983.

Chassin, M.R., Brook, R.H., Park, R.E., et al.: Variations in the Use of the Medical and Surgical Services by the Medicare Population. N Eng J Med 314 (5): 285-290, January 30, 1986.

Coffey, R.: Nonsurgical DRGs: Should They Be Refined with Disease Staging? Paper presented at the Second Annual Meetings of the Association for Health Services Research, Chicago, Ill., June 11, 1985.

Coffey, R., and Goldfarb, M.G.: DRGs and Disease Staging for Reimbursing Medicare Patients. Working Paper Number 1, Hospital Studies Program. Rockville, MD. National Center for Health Services Research, October, 1984.

Commission on Professional Hospital Activities: ICD-9-CM: International Classification of Diseases, 9th Revision, Clinical Modification. Vols. 1-3. Ann Arbor, Mich. Edwards Bros., 1978.

Conklin, J.E.: DRG Refinement: A Study of Alternative Groupings with Six Sets of Adjacent DRGs. Final Report on Subcontract 85-19 to HCFA cooperative agreement No. 18-C-98489/9-01. Santa Barbara, Calif. SysteMetics, Inc., March 1985.

Conklin, J.E., Lieberman, J.V., Barnes, C.A., et al.: Disease Staging: Implications for hospital reimbursment and management. *Health Care Financing Review*, 1984 Annual Supplement, November 1984b.

Conklin, J.E., Louis, D.Z., Lieberman, J.V., et al.: *DRG Refinement: A Feasibility Assessment Using Stage of Disease, Age and Unrelated Comorbidity*. Final Report on Contract No. HHS-100-82-0038. Santa Barbara, Calif. SysteMetrics, Inc., November, 1984a.

Cotterill, P., Bobula, J., and Connerton, R.: *A Comparison of Alternative DRG Relative Weights For The Medicare Prospective Payment System*. Working Paper. Baltimore, Md. Health Care Financing Administration, February 1985.

Coulton, C.J., McClish, S., Doremus, H., et al.: Implications of DRG Payments for Medical Intensive Care. *Med Care* 23(8): 977-985, August 1985.

Cretin, S., Newhouse, J.P., Buchanan, J., et al.: *Accounting Profits by Diagnosis Related Group and by Hospital Type Under Alternative Prospective Payment Methods*. Working Draft WD-2940-HCFA. Santa Monica, Calif. Rand Corp., February, 1986. *

Cretin, S., and Worthman, L.: *Alternative Systems for Case Mix Classification*. Working Draft WD-2927-HCFA. Santa Monica, Calif. Rand Corp., February, 1986. *

Cromwell, J., Mitchell, J.B., Calore, K.A., et al.: *Sources of Hospital Cost Variation by Urban-Rural Location, Teaching Status, and Bedsize*. Health Care Financing Administration Co-operative Agreement No. 18-C-98526/1-02. Waltham, Mass. Health Policy Research Consortium - Cooperative Research Center, December 26, 1985. *

Davis, C.K.,: Interhospital Differences in the Severity of Illness (letter). *N Eng J Med* *313* (18): 1163, October 31, 1985.

Demlo, L.K., and Campbell, P.M.: Improving Hospital Discharge Data: Lessons From the National Hospital Discharge Study. *Med Care* *19*(1): 1030-1040, October 1981.

Eggers, P.: Unpublished data. Baltimore, Md. Health Care Financing Administration, 1986.

Federal Register 48 (171):39752, September 1, 1983.

Federal Register 49 (171):34728, August 31, 1984.

Federal Register 50 (170):35646, September 3, 1985.

Fetter, R.B., Shin, Y., Freeman, J.L., et al.: Case Mix Definition by Diagnosis Related Groups. *Med Care* 18(Supplement): 1-53, February 1980.

Fetter, R.B., Thompson, J.D., Averill, R.F., et al.: The New ICD-9-CM Diagnosis Related Groups (DRG) Classification Scheme. Final Report to HCFA on Grant Nos. 95P 97499/1-01 and 95P 97499/1-02. New Haven, Conn. Yale University, May 1982.

Frank, R.G. and Lave, J.R.: The Psychiatric DRGs: Are They Different? Med Care 23(10): 1148-1155, October 1985.

GAO: Use of Unaudited Hospital Cost Data Resulted in Overstatement of Medicare's Prospective Payment Rates: Report to the Congress by the Comptroller General of the United States. GAO/HRD 85-74. Washington, D.C. General Accounting Office, July 18, 1985.

Garber, A.M., Fuchs, V.R., and Silverman, J.F.: Case Mix, Costs and Outcomes: Differences between Faculty and Community Services in a University Hospital. N Eng J Med 310(19): 1231-1237, May 10, 1984.

Gertman, P.M., and Bucher, B.M.: Inappropriate Hospital Bed Days and Their Relationship to Length of Stay Parameters. Paper presented at Annual Meeting of the American Public Health Association, October 1971.

Gertman, P.M., and Lowenstein, S.: A Research Paradigm for Severity of Illness: Issues for the Diagnosis-Related Group System. Health Care Financing Review 1984 Annual Supplement, November 1984.

Gertman, P.M., and Restuccia, J.D.: The Appropriateness Evaluation Protocol: A Technique for Assessing Unnecessary Days of Hospital Care. Med Care 19: 855, 1981.

Gonnella, J.S., (Ed.): Clinical Criteria for Disease Staging. Santa Barbara, Calif. SysteMetrics, 1983.

Gonnella, J.S., Hornbrook, M.C., and Louis, D.Z.: Staging of Disease: A Case-Mix Measurement. JAMA 251(5):637-644, February 3, 1984.

Gordon, N.M.: Testimony before the Subcommitte on Health, Committee on Finance, United States Senate. July 29, 1985.

Gornick, M.: Trends and Regional Variations in Hospital Use under Medicare. Health Care Financing Review. Vol. 3, No. 3. HCFA Pub. No. 03141. Office of Research and Demonstrations, Health Care Financing Administration. Washington, D.C. U.S. Government Printing Office, March 1982.

Guterman, S.: A Descriptive Analysis of PPS Outlier Payment Policy. Working Paper. Baltimore, MD. Health Care Financing Administration, March, 1986. *

HHS: Report to the Congress: The Impact of the Medicare Hospital Prospective Payment System, 1984 Annual Report. Washington, D.C. Department of Health and Human Services, 1985.

Horn, S.D.: The Computerized Severity Index: Conceptualization, Comparisons, and Applications. Baltimore, Md. Center for Hospital Finance and Management--The Johns Hopkins University, February, 1986.

Horn, S.D.: Hospital Planning for Profit: The Importance of Measuring Severity of Illness. Ala J Med Sci 22(1): 21-29, January 22, 1985.

Horn, S.D., Bulkley, G., and Horn, R.A.: Interhospital Differences in the Severity of Illness (letter). N Eng J Med 313 (18): 1164, October 31, 1985c.

Horn, S.D., Bulkley, G., Sharkey, P.D., et al.: Interhospital Differences in Severity of Illness. N Eng J Med 313(1): 20-24, 1985a.

Horn, S.D., Chacich, B., and Clopton, C.: Measuring Severity of Illness: A Reliability Study. Med Care 21(7): 705-714, 1983b.

Horn, S.D., and Horn, R.A.: Reliability and Validity of the Severity of Illness Index. Med Care (In Press, 1986).

Horn, S.D., and Horn, R.A.: A Study of Reliability of the Severity of Illness Index. Mimeo, preliminary Report, February, 1985.

Horn, S.D., Horn, R.A., and Sharkey, P.D.: The Severity of Illness Index as a Severity Adjustment to Diagnosis-Related Groups. Health Care Financing Review 1984 Annual Supplement, November 1984a.

Horn, S.D., and Sharkey, P.D.: Measuring Severity of Illness to Predict Patient Resource Use Within DRGs. Inquiry 20: 314-321, Winter 1983.

Horn, S.D., Sharkey, P.D., and Bertram, D.A.: Measuring Severity of Illness: Homogeneous Case Mix Groups. Med Care 21(1): 14-30, 1983a.

Horn, S.D., Sharkey, P.D., Chambers, A.F., et al.: Severity of Illness within DRGs. Final Report on HCFA grant 18-P-98378/3-01, December 1984b.

Horn, S.D., Sharkey, P.D., Chambers, A.F., et al.: Severity of Illness within DRGs: Impact on Prospective Payment. American Journal of Public Health 75(10): 1195-1199, 1985b.

Iezzoni, L.I., and Moskowitz, M.A.: The Clinical Impact of DRG-Based Physician Reimbursement. Waltham Mass. Health Policy Research Consortium, December 31, 1984.

Iezzoni, L.I., and Moskowitz, M.A.: Clinical Overlap Among Medical Diagnosis-Related Groups. JAMA 255(7): 927-929, February 21, 1986.

Institute of Medicine: Reliability of Medicare Hospital Discharge Records. Washington, D.C. National Academy of Sciences, November, 1977.

Institute of Medicine. Reliability of National Hospital Discharge Survey Data. Washington, D.C. National Academy of Sciences, 1980.

Institute of Medicine: Report of a Study: Reliability of Hospital Discharge Abstracts. Washington, D.C. National Academy of Sciences, February, 1977.

Jencks, S.F., and Kay, T.L.: Do Frail, Disabled, Poor and Very Old Medicare Beneficiaries Have Higher Hospital Charges? Working Paper. Baltimore, Md. Health Care Financing Administration, 1986. *

Johnson, A.N., and Appel, G.L.: DRGs and Hospital Case Records: Implications for Medicare Case Mix Accuracy. Inquiry 21: 128-134, Summer, 1984.

Kay, T.L., Jencks, S., and Terrell, S.: The Effect of Removing Nonacute Care on DRG Homogeneity. Working Paper. Baltimore, Md. Health Care Financing Administration, 1986. *

Kessner, D.M., Kalk, C.E., and Singer, J.: Assessing Health Quality--The Case for Tracers. N Eng J Med 288: 189-194, January 25, 1973.

Knaus, W.A., Draper, E.A., Wagner, D.P., et al.: Apache II: A Severity of Disease Classification System. Critical Care Medicine, 13(10): 818-29, 1985.

Knaus, W.A., Zimmerman, J.E., Wagner, D.P., et al.: APACHE-- Acute Physiology and Chronic Health Evaluation: A Physiologically-Based Classification System. Critical Care Medicine 9: 591-597. August, 1981.

Langenbrunner, J.: Improving Case Mix Measurement in Speical Classes of Facilities. Working Paper. Baltimore, MD. Health Care Financing Administration, 1986. *

Lave, J.R.: Is Compression Occurring in DRG Prices? Inquiry 22: 142-147, Summer 1985.

List, N.D., Fronczak, N.E., Gottlieb, S.H., et al.: A Cross-National Study of Differences in Length of Stay of Patients with Cardiac Diagnoses. Med Care 21(5): 519-530, May, 1983.

Lloyd, S.S., and Rissing, J.P.: Physician and Coding Errors in Patient Records. JAMA 254(10): 1330-1336, Sept 12, 1985.

Louis, D.Z., et al.: Disease Staging: A Clinically Based Approach to Measurement of Disease Severity, Vol. 4: Staging Software User's Manual. NCHSR contract 233-78-3001 (part of the Hospital Cost and Utilization Project). Santa Barbara, Calif. SysteMetrics, Inc., 1983.

Louis, D.Z., Barnes, C., Jordan, N., et al.: Disease Staging: A Clinically Based Approach to Measurement of Disease Severity, Vol. 5: Reabstracting Study. NCHSR contract 233-78-3001 (part of the Hospital Cost and Utilization Project). Santa Barbara, Calif. SysteMetrics Inc., 1984.

Luft, H.: How Do Health Maintenance Organizations Achieve Their "Savings": Rhetoric and Evidence. New Eng J Med 298(24):1336-1343, June 15, 1978.

Luft, H.: Health Maintenance Organizations: Dimensions of Performance. New York. John Wiley & Sons, 1981.

Manning, W.G., Lebowitz, A., Goldberg, G.G., et al.; A controlled Trial of the Effect of a Prepaid Group Practice on Use of Services. New Eng J Med 310(23):1505-1530, June 7, 1984.

McGuire, T.G.: Criteria for Evaluating Case-Mix Classification Systems as a Basis for Prospective Payment: An Application to Psychiatric Care. Unpublished. Waltham, Mass., Brandeis University, January, 1985.

McKibbin, R.C., Brimmer, P.F., Clinton, J.F., et al.: DRGs and Nursing Care. Final Report, HCFA Grant 15-C-98421/7-02. St. Louis Mo. American Nurses Association, 1985.

McNeil, B.J., Pauker, S.G., Sox, H.C., et al.: On the Elicitation of Preferences for Alternative Therapies. N Eng J Med 306:1259-1262, May 27, 1982.

McPherson, K., et al.: Small Variations in the Use of Common Surgical Procedures. N Eng J Med 307:1310-1314, November 18, 1982.

MediQual Systems Inc.: The Medical Illness Severity Grouping System (MEDISGRPS). Brochure. Chicago, Ill. MediQual Systems, Inc., 1984.

Mitchell, J.B., Calore, K.A., Cromwell, J., et al.: Creating DRG-Based Physician Reimbursement Schemes: A Conceptual and Empirical Analysis. First year Report to HCFA on Grant Number 18-P-98387/1-01. Chestnut Hill, Mass. Health Economics Research, October, 1984.

Mitchell, J.B., Calore, K.A., Cromwell, J., et al.: Physician DRGs: What Do They Look Like and How Would They Work? Draft Discussion Paper, HCFA Grant No. 18-P-98387/1-01. Chestnut Hill, Mass. Health Economics Research, February 25, 1985a.

Mitchell, J.B., Calore, K.A., Cromwell, J., et al.: The Role of Physician Practice Patterns in within-DRG Cost Variation. Health Care Financing Administration Cooperative Agreement No. 18-C-98526/1-02. Waltham, Mass. Health Policy Research Consortium-Cooperative Research Center, December 19, 1985b. *

Morrison, L.J., Freiman, M.P., and Dickey, B.: A Study of Patient Classification Systems for Prospective Rate-Setting for Medicare Patients in General Psychiatric Units and Psychiatric Hospitals. Final Report, NIMH contract No. 278-84-0011(DB). Rockville, MD. National Institute of Mental Health, December 31, 1985.

Mullin, R.L.: Interhospital Differences in the Severity of Illness (letter). N Eng J Med 313 (18) 1164, October 31, 1985.

Munoz, E., Laughlin, A., Regan, D.M., et al.: The Financial Effects of Emergency Department-Generated Admissions Under Prospective Payment Systems. JAMA 254(13):1763-1771, October 4, 1985.

Munoz, E., Soldano, R., Laughlin, A., et al.: Source of Admission and Cost: Public Hospitals Face Financial Risk. Am J. Pub. Hlth 76(6):696-697, June 1986.

Nathanson, M.: More Hospitals Turn to SOIs, but Experts Question Their Usefulness. Modern Healthcare 15(4):63-75, February 15, 1985.

National Associations of Children's Hospital and Related Institutions: Children's Hospital's Case Mix Classification System Project, Phase 1 Report: Critique of DRGs Across Hospital Settings. Prepared by Policy Analysis, Inc. for NACHRI under Health Care Financing Administration Grant No. 95-P-98570/3-01. Alexandria, Va. National Association of Children's Hospitals and Related Institutions, Inc., March 1, 1985.

National Associations of Children's Hospital and Related Institutions: Children's Hospital's Case Mix Classification Project, Phase 2 Final Report. Prepared by Policy Analysis, Inc. for NACHRI under Health Care Financing Administration Grant No. 95-P-98570/3-01. Alexandria, Va. National Association of Children's Hospitals and Related Institutions, Inc., October 24, 1985.

Norbrega, F.T., Naessens, J.M., Kelly, M.P., et al.: Feasibility of Using CPT-4 Codes for DRG Assignment of Surgical Patients. Final Report, HCFA Cooperative Agreement No. 18-C-98631/5-01. Rochester, Minn. Mayo Clinic, November, 1985.

OIG: Memo from R.K. Kusserow, Inspector General, Department of Health and Human Services, to C. McClain Haddow, Acting Administrator, Health Care Finance Administration re Priority Audit Memorandum--Large Profits Earned by Hospitals under the Medicare Prospective Payment System. ACN: 09-62021, Washington, D. C. Department of Health and Human Services, October 29, 1985.

Pauly, M.V., and Wilson, P.: The Marginal Cost of Hospital Output and Empty Beds. Final Report pursuant to NCHSRHCTA Grant No. HSO4618. Lansing, Mich. Michigan Hospital Association, May 1985.

Pettengill, J., and Vertrees, J.: Reliability and Validity in Hospital Case-Mix Measurement. Health Care Financing Review 4(2):101-128, December 1982.

Restuccia, J.D., Gertman, P.M., et al.: A Comparative Analysis of Appropriateness of Hospital Use. Health Affairs 3(2):130-138, Summer 1984.

Richards, T., Rogers, W., Reboussin, D., et al.: Reliability Analysis of the Severity of Illness Index: Preliminary Results Based on Data from Five Hospitals. Working Draft, WD-2892-HHS, Santa Monica, Calif., Rand Corp., 1986.

Sheingold, S.: The Indirect Teaching Adjustment for Medicare's Prospective Payment System: Issues and Options. Washington, DC. Congressional Budget Office, April, 1985.

Short, T.L., and Coffey, R.M.: Diagnosis Related Groups vs. Disease Staging: Implications for Hospital Reimbursement. Hospital Studies Program Working Paper No. 5. Rockville, Md. National Center for Health Services Research, October 1984.

Simborg, D.W.: DRG Creep: A New Hospital-Acquired Disease. N Eng J Med 304:1602-1604, June 25, 1981.

Stern, R.S., and Epstein, A.M.: Institutional Responses to Prospective Payment Based on Diagnosis-Related Groups. N Eng J Med 312(10):621-627, March 7, 1985.

Strumwasser, I., and Paranjpe, N.: The Appropriateness of Hospitalization: A Comparative Analysis of the Appropriateness Evaluation Protocol and the Standardized Medreview Instrument. Detroit, Mich. Michigan Health Care Education and Research Foundation, Inc., June 30, 1985.

SysteMetrics, Inc.: The Measurement of Inappropriateness Hospital Utilization in a Nationally Representative Sample. Santa Barbara, California: SysteMetrics, Inc., January 1983.

Thompson, J.D.: The Measurement of Nursing Intensity. Health Care Financing Review 1984 Annual Supplement, November 1984.

Thorpe, K.E., Cretin, S., and Keeler, E.B.: Are the DRG Case Weights Compressed? Working Draft WD-2606-HCFA-1. Santa Monica, Calif. Rand Corp., February, 1986. *

Wagner, D.P., and Draper, E.A.: Acute Physiology and Chronic Health Evaluation (APACHE II) and Medicare Reimbursement. Health Care Financing Review 1984 Annual Supplement, November 1984.

Wagner, D.P., Draper, E.A., Campos, R.A., et al.: Initial International Use of APACHE: An Acute Severity of Disease Measure. Medical Decision Making 4(3):297-313, 1984.

Wennberg, J., and Gittelsohn, A.: Variations in Medical Care Among Small Areas. Scientific American 246(4):120-134, 1982.

Wennberg, J.E., McPherson, K., and Caper, P.: Will Payment Based On Diagnosis-Related Groups Control Hospital Costs? N Eng J Med 311(5):295-300, August 2, 1984.

West, H., Marcus, L., McMenamin, P., et al.: Physician and Hospital Reimbursement Study. Report, HHS Contract No. 100-83-0026. Vienna, Va. Mandex, Inc., January 31, 1985.

Worthman, L.G., and Cretin, S.: Review of Literature on Diagnosis Related Groups. Working Draft WD-2316-1-HCFA. Santa Monica, Calif. Rand Corp., March, 1986.

Young, W.W.: Incorporating Severity of Illness and Comorbidity in Case-mix Measurement. Health Care Financing Review, 1984 Annual Supplement, November, 1984.

Young, W.W.: Measuring the Cost of Care Using Patient Management Categories. Final Report, HCFA grant 18-P-97063/3-05, (Vol I) January 17, 1985.

Young, W.W., Joyce, D.Z., Wolrath, J.M., et al.: Effectiveness of Two Methods of Patient Classification. Appendix 3.2 in Young W.W., Measuring the Cost of Care Using Patient Management Categories, Final Report, 1985.

Young, W.W., Swinkola, R.B., and Zorn, D.M.: The Measurement of Hospital Case Mix. Med Care, 20(5):501-512, 1982.

*U. S. GOVERNMENT PRINTING OFFICE 1987: 193-160/70309

CMS LIBRARY
3 8095 00012306 3

U.S. Department of
Health and Human Services
Health Care Financing Administration
1A9 Oak Meadows Building
6325 Security Boulevard
Baltimore, MD 21207

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Dept. of H.H.S.
HHS-392

SPECIAL FOURTH-CLASS RATE

U.S.MAIL

U.S. Department of Health and Human Services
Health Care Financing Administration
Bureau of Data Management and Strategy
HCFA Pub. No. 03254 September 1987